Projective Minimal Analysis of Camera Geometry

by

Raquel A. Romano

A.B. Mathematics, Harvard University, 1992
S.M. Electrical Engineering and Computer Science, M.I.T., 1995

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2002

Author .....
Department of Electrical Engineering and Computer Science
May 24, 2002

Certified by .....
W. Eric L. Grimson
Professor of Computer Science and Engineering
Thesis Supervisor

Certified by .....
Olivier D. Faugeras
Adjunct Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Projective Minimal Analysis of Camera Geometry

by

Raquel A. Romano

Submitted to the Department of Electrical Engineering and Computer Science
on May 24, 2002, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

This thesis addresses the general problem of how to find globally consistent and accurate estimates of multiple-view camera geometry from uncalibrated imagery of an extended scene. After decades of study, the classic problem of recovering camera motion from image correspondences remains an active area of research. This is due to the practical difficulties of estimating many interacting camera parameters under a variety of unknown imaging conditions. Projective geometry offers a useful framework for analyzing uncalibrated imagery. However, the associated multilinear models—the fundamental matrix and trifocal tensor—are redundant in that they allow a camera configuration to vary along many more degrees of freedom than are geometrically admissible.

This thesis presents a novel, minimal projective model of uncalibrated view triplets in terms of the dependent epipolar geometries among view pairs. By explicitly modeling the trifocal constraints among projective bifocal parameters—the epipoles and epipolar collineations—this model guarantees a solution that lies in the valid space of projective camera configurations. We present a nonlinear incremental algorithm for fitting the trifocally constrained epipolar geometries to observed image point matches. The minimal trifocal model is a practical alternative to the trifocal tensor for commonly found image sequences in which the availability of matched point pairs varies widely among different view pairs.

Experimental results on synthetic and real image sequences with typical asymmetries in view overlap demonstrate the improved accuracy of the new trifocally constrained model. We provide an analysis of the objective function surface in the projective parameter space and examine cases in which the projective parameterization is sensitive to the Euclidean camera configuration. Finally, we present a new, numerically stable method for minimally parameterizing the epipolar geometry that gives improved estimates of minimal projective representations.

Thesis Supervisor: W. Eric L. Grimson
Title: Professor of Computer Science and Engineering

Thesis Supervisor: Olivier D. Faugeras
Title: Adjunct Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to thank my advisor, Eric Grimson, for his constant support throughout my graduate school years, for his valuable insights at the many stages of my research, and for his encouragement of all my intellectual pursuits. I would also like to thank my co-advisor Olivier Faugeras, for motivating my interest in projective geometry and for valuable conversations at the A.I. Lab and at INRIA. In addition, I would like to thank my readers, Tomas Lozano-Perez, and in particular Berthold Horn, for his key insights into the mathematical and practical issues regarding camera geometry.

I am grateful to the many characters of the A.I. Lab, past and present, whose curiosity and creativity make the lab a genuinely unique place to work, think, and play. Thanks to David Demirdjian for reading an early draft of this work and thanks especially to Jose Robles for reading countless drafts and revisions and offering his valuable technical insights at many stages of this work. Thank you to Polina Golland, Lily Lee, Erik Miller, Chris Stauffer, Kinh Tieu, and Lilla Zollei for helpful discussions, and also to Theo Papadopoulo for useful technical conversations during my stay at INRIA. In addition, thanks to many friends at the lab over the years for their support and entertainment at all hours of the day or night, especially Akin Aina, Christine Alvarado, Robbin Chapman, Jill Fekete, Polina Golland, Charles Isbell, Matt Marjanovic, Lauren O'Donnell, Jose Robles, and Lilla Zollei.

Finally, I am grateful to all my friends and family for their years of baffled patience while I completed this epoch of my education—to my older and younger brothers and their families, to my grandparents, and to the one and only Jesus Quintero for sharing with me his contagious joy for life. Most of all I would like to thank my amazing parents for believing in me.

This thesis is dedicated to my loving and lovable, fascinating and intellectually curious, late grandfather, Dr. Marvin D. Martin, who first inspired me to learn about computers.

Contents

Chapter 1

Introduction

Computer vision is the study of computational methods for understanding visual imagery. One of the earliest, most classic problems in computer vision is the question of how to compute the 3D positions and orientations of a set of cameras using a single 2D image taken from each camera. The classic approach to this problem is to measure points in each image that represent the same point in the world and express their 2D locations in terms of the unknown relative 3D camera poses. By modeling the physics and geometry of image formation, researchers were able to find equations relating observed image points to unknown camera and scene configurations. It was thought that the simple rearrangement of these equations would give rise to straightforward methods for computing the 3D poses of cameras. However, after several decades of research, the robust and accurate recovery of camera poses from matching points in image collections is still an active area of research.

Meanwhile, the motivating applications have changed dramatically. The traditional problems of egomotion recovery for mobile robotics and bundle adjustment for photogrammetry are now accompanied by a range of new applications due to the increased ease of acquiring, processing, and storing vast amounts of digital imagery and video. From artificial intelligence to computer graphics, there is a range of application areas that require accurate knowledge of the positions of cameras that have captured a set of possibly uncalibrated imagery. Activity monitoring, vision-enhanced vehicle guidance, robotic and vision-guided surgery, virtual world modeling, and 3D special effects in video are a few examples. While these applications may call for a wide variety of methods and models for interpreting infor-

mation about the scene being viewed, their performance often depends critically on having an accurate model of the camera configuration.

The primary difficulty facing 3D camera and scene analysis is uncertainty in the matching of image points across multiple views. Imaging noise, occlusions and disocclusions, appearance changes, and limited fields of view, cause many potential image matches to be unreliable. The fewer the number of reliably matched points viewed in two cameras, the more difficult it is to accurately compute the relative 3D camera positions. When a set of images covers an extended scene using traditional cameras, whether it be a collection of stationary cameras placed around a common scene or a single video camera capturing frames as it moves within an environment, many view pairs have only small regions of overlap in their fields of view. Furthermore, frames that do overlap may have markedly different views of the scene, making it difficult to locate matching scene points. This asymmetry in view overlap exacerbates the problem of computing a precise, globally consistent representation of a camera collection. This thesis addresses the problem of finding globally consistent relative pose estimates of uncalibrated cameras with sparsely overlapping fields of view.

1.1 General Thesis Contributions

The primary contribution of this thesis is a new approach to modeling and estimating the geometry of three uncalibrated views. We introduce a new projective parameterization of three views in general position that uses the trifocal lines to constrain all pairwise projective parameters. The parameterization is a *minimal* representation of an uncalibrated view triplet, meaning that the number of free parameters does not exceed the geometric model's physical degrees of freedom. From the geometric and algebraic constructions of the trifocally constrained model, we devise an algorithm for computing the minimal projective geometry of a view triplet from dependent pairwise projective relations. We show that this model is empirically valuable when there is an asymmetry in the availability of reliable point matches between the three view pairs, a common scenario in practical applications. The minimal parameterization guarantees that the model is theoretically consistent in that it corresponds to a unique projective camera configuration. Furthermore, the minimality of

the formulation is numerically favorable because it allows optimizations of the projective parameters to be done in a space of lowest possible dimensionality. Finally, the model may be fit to all the available data since it may be evaluated on either point matches across view pairs or across view triplets, a distinct advantage over trifocal models that are only associated with point triplets. The advantages of this approach are demonstrated on both synthetic and real image sets that exhibit common asymmetries in the availability of reliable point correspondences.

In addition, this thesis examines the relationship between the minimal parameters that define the projective geometry of view pairs and the shapes of the objective functions typically used to fit projective geometric models to image data. By analyzing the ambiguities in mappings between points in the projective parameter space and their Euclidean counterparts and considering their effective constraints on image point locations, we show that regions of uncertainty in projective parameter space have strong implications for the convergence behavior of nonlinear optimizations. In particular, we look at commonly encountered camera configurations that are nearly degenerate cases for the general camera model and propose a new method for choosing geometrically well-behaved minimal parameterizations that ensure desirable numerical behavior during optimization. We conclude by sketching the use of asymmetric minimal parameterizations for analyzing the geometry of arbitrarily many views.

1.2 The Basic Problem: From Images to Cameras

The input to the problem is a collection of 2D images of a scene viewed from different 3D positions. The 3D relationship between the set of scene points relative to the set of camera positions is assumed to be a rigid Euclidean transformation. Points are measured in multiple images by an image matching or tracking algorithm. The image point matcher determines where a scene point has moved from one frame to the other and returns the pair of 2D coordinates of the two matched image points, also called a *point correspondence*. Given many images, a point matching procedure may return pairs, triplets, or n-tuples of corresponding image points for any given scene point. Once the point matching algorithm

has computed the locations of matching image points, the remaining unknown values are the relative camera positions, the intrinsic camera properties, and the scene point locations. If the matched image points are treated as noisy measurements, the true 2D image projections of the viewed scene points may be added to the set of unknowns. Given a set of measured image point correspondences, the general problem of 3D camera analysis is to compute the 3D location and 3D orientation of each camera at the moment that its image was captured, also known as the *external* or *extrinsic* parameters, and the amount of scaling, skewing, and shifting that each individual camera applies to its image points, also known as the *internal* or *intrinsic* parameters. This work addresses the intermediate step of fitting a *projective* geometric model to uncalibrated imagery before proceeding to a Euclidean analysis of the internal and external parameters.

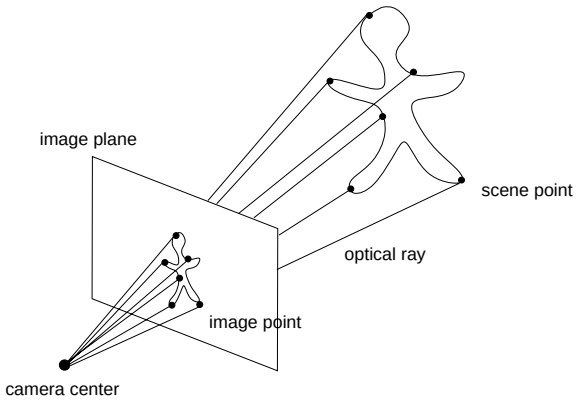


Figure 1-1: *The pinhole camera model for image formation.*

The basic geometry of imaging is traditionally modeled as a perspective projection via the pinhole camera model. An *optical ray* is defined as the ray originating at the camera's optical center, passing through the image plane at the 2D image point, and intersecting the 3D scene point (Figure 1-1). In the case of uncalibrated imagery, each image point undergoes an additional 2D affine transformation due to the intrinsic camera parameters. The transformation from scene points to image points can be modeled by a simple set of equations involving the position of a 3D scene point, the external and internal camera parameters, and the 2D image point coordinates. When several cameras view the same scene

point, the perspective projection equations are all coupled through the 3D coordinates of the scene point. These equations are useful for recovering unknown quantities from known quantities. If the camera parameters and scene point positions are already known, then the equations dictate exactly where the 2D matching image points must be. Conversely, in the classical problem of 3D camera analysis the positions of matching image points are known, and the equations provide constraints on the 3D configuration of the scene points and cameras.

Historically, the problem of recovering camera geometry, or *motion*, has been coupled with that of finding the 3D positions of scene points, or *structure*. This problem has been studied in the photogrammetry domain [4] and in computer vision as the *structure from motion* problem [30, 50]. This thesis focuses on the problem of estimating relative camera geometry from uncalibrated images, a task that provides a necessary starting point for many structure from motion procedures.

To estimate camera geometry directly from point matches, we make use of relations that constrain matching image points in terms of the camera poses. These constraints are naturally expressed in terms of *relative* camera poses. Even the known multiple-view constraints on n-tuples of points—the trifocal, quadrifocal, and multilinear tensors—only encode relative relations between selected view pairs. Only when an estimate of 3D scene structure is known can the scene points be used to define a global coordinate system that couples all cameras simultaneously. However, to estimate scene structure, a fairly accurate initial estimate of relative camera positions is necessary. This is the focus of our work.

1.3 Projective Minimal Models

This thesis addresses several related difficulties in 3D camera analysis. The first is the problem of finding a globally consistent model of multiple cameras. Since the camera information directly recoverable from image points expresses pairwise relations, it is possible to estimate each camera pair's relative pose independently of all others by using only matching points between pairs of views. The danger of this strategy, as depicted in Figure 1-2, is that three independent camera pair relations are not necessarily consistent with a meaningful

global camera configuration.

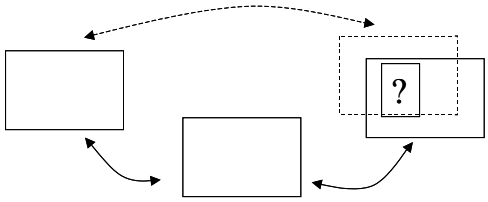


Figure 1-2: *Three independent pairwise camera relations are not necessarily consistent with a single global camera configuration.*

When the physical cameras are available for internal calibration by positioning a known 3D calibration pattern in each camera's field of view, the task may be reduced to a Euclidean 3D camera analysis problem known as *multiple baseline stereo*. While still a difficult problem, global consistency may be enforced via the transitive properties of 3D translations and rotations. When the original cameras are not available, and the calibration information is unknown, as is commonly the case when analyzing an arbitrary collection of images, Euclidean consistency may not be enforced.

When the image collection is not calibrated, the projective framework offers a powerful machinery for analyzing 3D camera positions. In this framework, constraints on matching image points are expressed as projective relations among pairwise cameras. These relations are expressed in terms of geometric structures that couple the external and internal camera parameters and carry valuable information about the Euclidean configuration. Though powerful, the projective setting presents a second difficulty that this thesis addresses, that of finding a globally consistent *projective* representation of the camera geometry. To be globally consistent, a projective model must have only as many independent parameters as are allowable by the degrees of freedom allowable in the system. We call such models *minimal* projective representations.

The final difficulty addressed by this thesis is a common practical scenario when many cameras view an extended scene. Due to occlusions, noise, and the limited fields of view of typical cameras, not all cameras will view the same set of scene points. While those camera pairs with significant view overlap will provide an image matcher with many easily

matched points, other pairs with less overlap will have only a few good point correspondences (Figure 1-3). Furthermore, for three or more views it is often much harder to find triplets or n-tuples of points, than it is to find matched point pairs across two views at a time. This practical issue of asymmetric view overlap makes it difficult to use matching image points between all view pairs to reliably estimate the relative positions of all cameras.

Figure 1-3: *The number and reliability of point matches between some view pairs is greater than between other view pairs. The two consecutive view pairs have greater view overlap than the non-consecutive pair.*

1.4 View-Based Constraints

Throughout this thesis, we make heavy use of view-based constraints on matching image points, that is, constraints that involve only image points and camera positions, not 3D scene points. In the pinhole camera model, each optical ray contains the camera center, the image point, and the scene point (Figure 1-1). The basic constraint on two views of a scene point is that two optical rays through matching image points must intersect at a single scene point. This relation may be simply stated by requiring that the two optical rays be coplanar. If each optical ray is expressed as the ray joining the unknown camera center with the known 2D image point, then the so-called *coplanarity constraint* guarantees that they intersect at single 3D scene point (Figure 1-4). The coplanarity constraints on all pairs of matching points between two views give rise to a set of equations that involve only the relative camera positions and image point coordinates, without explicitly involving the 3D scene point.

For three views of a scene point, all three optical rays must intersect at a single 3D

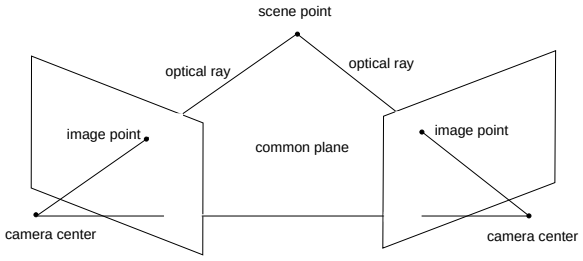


Figure 1-4: *Two optical rays through matching image points must be coplanar.*

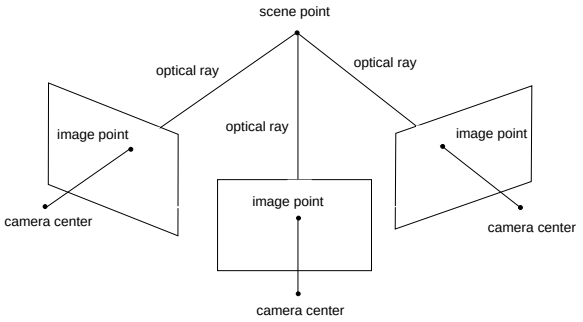


Figure 1-5: *Three optical rays through matching image points must intersect at a single scene point.*

point (Figure 1-5). Again, this may be written as a view-based constraint that does not involve the scene point. The advantage of these view-based expressions of optical ray intersections is that the unknown scene point locations are eliminated from the equations and therefore estimates of camera positions may be separated from estimates of scene point depths. These expressions give rise to error functions that evaluate how well a model of the camera geometry explains the measured image matches. Finally, since these expressions involve only incidence relations among points, lines, and planes, they are defined in both the Euclidean framework traditionally used for calibrated imagery, and in the projective framework that we will use for analyzing uncalibrated imagery.

1.5 Projective Representations

In a typical imaging device, the properties of the camera lens and the imaging plane or CCD cause an effective offset, scaling, and shearing transformation of the image points after perspective projection. When the intrinsic camera properties are known, this affine transformation may be inverted to rectify the image points to a scene-based coordinate frame. This is called the calibrated case, and it permits the imaging process to be modeled in a Euclidean framework. When the intrinsic properties are not known, the Euclidean model is not applicable. In the projective framework, the perspective projection of 3D scene points composed with the 2D affine transform of their 2D projections are treated as a single projection of the 3D projective space onto the 2D projective plane. This embedding of the Euclidean space containing cameras, scene points, and image points into a projective space preserves incidence relations among points, lines, and planes. These incidence relations allow the view-based constraints on image points to be expressed without explicit knowledge of the internal parameters.

In the projective setting, camera relations are described by parameters that depend jointly on the unknown internal parameters and on the 3D external rotations and translations. However, the projective constraints on observed image points may be defined entirely in terms of the 2D images. These projective parameters—the epipoles and the transformation of the epipolar pencil—directly constrain the locations of matching image points, so they may be computed from measured correspondences. Once the projective relations among cameras are established, it is often possible to recover the internal parameters without knowledge of a calibration pattern [10, 51, 26]. Using the internal parameters, a projective representation may be projected back into the Euclidean setting.

The accurate estimation of projective parameters is thus a useful first step in the analysis of uncalibrated imagery. However, the projective representation allows certain freedoms in the camera model that are not physically realizable because they do not enforce metric properties such as distances and angles. While a Euclidean camera and scene model is defined up to an arbitrary 3D rigid transformation, a projective camera model is defined up to an arbitrary 3D projective transformation. The greater the error in image point measurements, the more the accuracy of a projective model is likely to be corrupted, as for

any over-parameterized model.

This thesis aims to exploit the usefulness of projective relations while enforcing consistency within the projective framework in order to overcome practical issues of view overlap asymmetry in image sequences.

1.6 Specific Thesis Contributions

This work addresses a fundamental problem in 3D camera analysis: given multiple uncalibrated views of a scene, compute an accurate estimate of the relative camera geometry, for delivery to a bundle adjustment or structure from motion procedure, or any other system requiring initial knowledge of camera locations. We present a new theoretical framework for using epipolar geometry estimates between certain view pairs to constrain the relations between other view pairs. The model guarantees a minimal parameterization of a view triplet, which is the minimal set of views required to perform self-calibration, the unsupervised recovery of intrinsic camera properties relations [10], and the relative scaling of camera baselines.

We present a new estimation procedure for the model and give experimental evidence of its improved accuracy for camera pose estimation in practical imaging scenarios with asymmetric view overlap. The minimal trifocal model may be used in a wider context to perform projective bundle adjustment or to interleave camera estimation with scene point estimation or image point re-estimation.

We analyze the objective function defined on the minimal parameter space for two or three views and examine its behavior on image point matches. We examine sensitive camera configurations and show the existence of numerically unstable regions in the parameter space. We present a novel generalized parameterization of the two-view and three-view projective relations that avoids numerical instabilities and present empirical evidence of its accuracy for estimation.

The theoretical impact of this work is that it models three views in general position using the minimal allowable degrees of freedom and view-based, projective constraints. Estimates of the trifocally constrained epipolar parameters are guaranteed to satisfy all geometric

dependencies and therefore correspond to an admissible projective camera configuration.

1.7 Thesis Outline

Chapter 2 introduces the projective camera model, derives the known projective constraints on two, three, or arbitrarily many views, describes the relationship between multilinear tensors and minimal parameter sets, and reviews related work on fitting projective camera models to image collections.

Chapter 3 presents the new minimal parameterization of three views. It describes the geometric dependencies among epipolar parameters, namely the restriction of epipoles to trifocal lines and the fixed point of the epipolar collineation given by the trifocal line. Using the key observation that trifocal lines may be expressed using only knowledge of the epipolar geometry between two view pairs, we proceed to construct a trifocally constrained minimal model of the epipolar relations among all view pairs in a triplet.

Chapter 4 presents a procedure for fitting the minimal model from Chapter 3 to matching points in a view triplet via nonlinear optimization. It presents experimental results using synthetic and real image sequences with asymmetric overlaps in their fields of view. The minimally parameterized estimation procedure is shown to improve the accuracy of the epipolar geometry between two views with very few reliable point matches.

Chapter 5 examines the shapes of the objective function for the epipolar parameters, revealing regions of uncertainty in the projective parameter space that can potentially trap searches. Experimental results on both synthetic and real imagery are presented, showing how false minima are more likely to be encountered for certain camera configurations. Finally, this chapter presents a generalization of the epipolar geometry parameterization that avoids degeneracies by automatically choosing a stable parameterization of the epipolar transformation.

Chapter 6 discusses issues of robustness, degeneracy, and the applicability of this work to arbitrarily long sequences.

Chapter 2

Multiple View Geometry

This chapter provides the theoretical background for modeling multiple views of a scene in the framework of projective geometry. Starting from the pinhole camera model, we describe the coplanarity constraints between two uncalibrated views in terms of the epipoles and the transformation of the epipolar pencil, also called the epipolar collineation. We derive a known minimal parameterization of the epipolar geometry and its relation to the fundamental matrix [5, 53, 9]. We refer to the resulting minimal parameter set as the *epipolar parameters*. These form the building block of the remainder of the thesis. We briefly describe the trifocal tensor that constrains matching point triplets and its generalization to the multilinear tensors for arbitrarily many views. We conclude by reviewing existing methods for camera and scene understanding using projective analysis.

2.1 One View

2.1.1 Camera Model

Throughout most of this work, the perspective projection of a 3D scene point onto a 2D image plane is modeled using projective geometry and the pinhole camera model. An image plane is represented by the projective plane $\mathbb{P}^2$, defined as the real Euclidean plane $\mathbb{R}^2$ augmented by a line at infinity. The 3D world is represented by the three-dimensional projective space $\mathbb{P}^3$, defined as the real Euclidean space $\mathbb{R}^3$ augmented by a plane at infinity. For a thorough exposition on projective geometry applied to computer vision, see [8, 21].

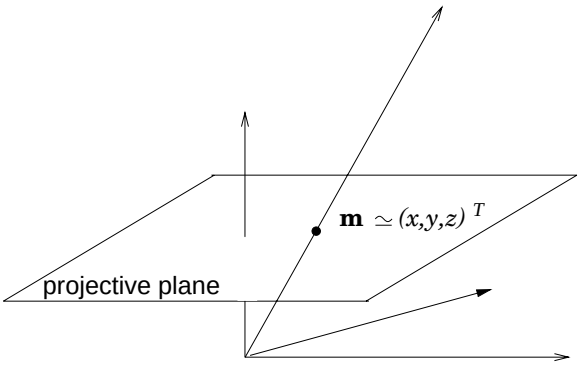


Figure 2-1: *The projective plane* $\mathbb{P}^2$ *embedded in* $\mathbb{R}^3$. *A point* $\mathbf{m}$ *in the projective plane is represented in homogeneous coordinates by any nonzero scalar multiple of the vector* $(x, y, z)^T$.

Points in a projective space are represented by *homogeneous coordinates*, that is, they are defined only up to a nonzero scale factor. An intuitive way to understand this representation is to consider the two-dimensional case of the projective plane. Figure 2-1 depicts the projective plane $\mathbb{P}^2$ embedded in $\mathbb{R}^3$. The point $\mathbf{m}_{\mathbf{P}^2}$ is associated with the set of all scalar multiples of the real vector $\mathbf{m}_{\mathbf{R}^3} = (x, y, z)^T$. All the members of this family intersect the projective plane at a single point, so we write the point in homogeneous coordinates as $\mathbf{m}_{\mathbf{P}^2} \simeq (x, y, z)^T$, where the notation $\simeq$ implies equality up to scale. The same applies to points in $\mathbb{P}^3$: a scene point is written in homogeneous coordinates as $\mathbf{M}_{\mathbf{P}^3} \simeq (X, Y, Z, W)^T$.

Given a 2D or 3D point defined in Euclidean coordinates, we embed it in a projective space by the simple mapping,

$$\mathbf{m}_{\mathbf{R}^2} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \rightarrow \quad \mathbf{m}_{\mathbf{P}^2} \simeq \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{M}_{\mathbf{R}^3} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad \rightarrow \quad \mathbf{M}_{\mathbf{P}^3} \simeq \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

The subscripts denoting real or projective coordinates are omitted throughout the rest of

this work, and quantities defined by $\simeq$ are assumed to be projective. A point whose last coordinate is equal to 0 is a point at infinity in the projective space, e.g. $(x, y, 0)^T \in \mathbb{P}^2$ and $(X, Y, Z, 0)^T \in \mathbb{P}^3$ are points at infinity. The origin of $\mathbb{P}^2$ is $(0, 0, 1)^T$ and the origin of $\mathbb{P}^3$ is $(0, 0, 0, 1)^T$.

The pinhole camera model projects a scene point $\mathbf{M} \in \mathbb{P}^3$ through the camera's optical center $\mathcal{C}$ onto the image point $\mathbf{m} \in \mathbb{P}^2$. The perspective projection is expressed as a linear projective transformation by the 3×4 camera matrix $\mathcal{P}$:

$$\mathbf{m} \simeq \mathcal{P}\mathbf{M}. \tag{2.1}$$

Any matrix representing a projective linear transformation is also defined only up to a scale factor since the image of a projective point under the scaled transformation leaves its homogeneous coordinates unchanged. It is often convenient to partition the camera matrix as $\mathcal{P} \simeq [\mathbf{P} \;\; \mathbf{p}]$, where $\mathbf{P}$ is a 3×3 matrix and $\mathbf{p}$ is a 3-element column vector. The next section describes the form that a matrix $\mathcal{P}$ must take to represent a physical camera. The optical center of the camera is defined as $\mathcal{C} \in \mathbb{P}^3$ such that $\mathcal{P}\mathcal{C} = \mathbf{0}$. We assume throughout that the optical centers of the cameras are not at infinity. It can be shown [8] that under this assumption, $\mathbf{P}$ is nonsingular, and the optical center has the form

$$\mathcal{C} \;\; \simeq \;\; [-\mathbf{P}^{-1}\mathbf{p}, \;\; 1]^T. \tag{2.2}$$

A projective line in $\mathbb{P}^2$ is represented by a 3-vector whose components are the coefficients of the linear form defining it. For example, the projective line $\mathbf{l}$ that contains all projective points $\mathbf{m} \simeq (x, y, z)^T$ such that $ax + by + cz = 0$ is represented in homogeneous coordinates as $\mathbf{l} \simeq (a, b, c)^T$. The equation of the line is $\mathbf{l}^T\mathbf{m} = 0$. Any nonzero multiple of $\mathbf{l}$ defines the same set of points and hence the same line. There is a dual relationship in $\mathbb{P}^2$ between points and lines.

Similarly, a projective plane in $\mathbb{P}^3$ is represented by a 4-vector representing its linear form. The plane $\mathbf{\Pi}$ containing all points $\mathbf{M} \simeq (X, Y, Z, W)^T$ such that $AX + BY + CZ + DW = 0$ is represented in homogeneous coordinates as $\mathbf{\Pi} \simeq (A, B, C, D)^T$, and the equation of the plane is $\mathbf{\Pi}^T\mathbf{M} = 0$. There is a dual relationship in $\mathbb{P}^3$ between points and planes.

The line $\mathbf{l}_\infty \simeq (0,0,1)^T$ is called the line at infinity and contains all the points at infinity in $\mathbb{P}^2$. The plane $\mathbf{\Pi}_\infty \simeq (0,0,0,1)^T$ is called the plane at infinity and contains all the points at infinity in $\mathbb{P}^3$.

2.1.2 Internal and External Parameters

The camera projection matrix can be factored into its intrinsic and extrinsic parameters as

$$\mathcal{P} \simeq \mathbf{A}[\mathbf{R}\ \mathbf{T}], \tag{2.3}$$

where the effect of a camera's intrinsic properties on the imaging process is modeled by the affine transformation,

$$\mathbf{A} \simeq \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.4}$$

The matrix $\mathbf{A}$ models 5 intrinsic parameters. The parameters f_x and f_y, the focal length of the camera measured in pixel units along the two coordinate axes of the camera's image plane. These may be distinct when a digital camera's vertical and horizontal sampling rates differ. The value s models the angle θ between the image plane's two coordinate axes by $s = \cos(\theta)$. The intersection of the camera's optical axis with the image plane is defined in Euclidean coordinates as the principal point, $(c_x, c_y)^T$, and typically lies near the center of the image. For many cameras, the focal lengths are equally scaled along both axes, so $f_x = f_y$ and the axes are orthogonal, so $s = 0$.[1] Many cameras often induce an additional radial distortion of the image that is nonlinear in the image coordinates which may be corrected [39, 49].

The Euclidean rotation and translation of the scene's global coordinate system relative to the camera's local coordinate frame are defined by a rotation matrix, $\mathbf{R}$, and a translation matrix, $\mathbf{T}$, each having 3 degrees of freedom. Therefore, every additional camera adds

[1] Although in this case there are 3 rather than 5 free internal parameters, the joint effects of the internal and external parameters on image points are not separately observable in uncalibrated imagery.

$5 + 3 + 3 = 11$ degrees of freedom to the configuration.

The transformation $\mathbf{m} \simeq \mathbf{A}[\mathbf{R} \;\; \mathbf{T}]\mathbf{M}$ from scene coordinates to image coordinates is thus decomposable into a 3D rigid Euclidean transformation followed by a nonrigid 2D transformation of the image plane. If a camera has been calibrated off-line by carefully measuring how a set of known 3D scene points are mapped to their 2D images, the inverse of the matrix $\mathbf{A}$ may be used to transform measured image points so that the only unknown parameters are the 3D rotations and translations. Unfortunately, this information is most often not available, as only the raw images or video are provided.

In the projective model, projection matrices are treated in their unfactored form, $\mathcal{P} \simeq [\mathbf{P} \quad \mathbf{p}]$. Each additional camera still contributes 11 degrees of freedom, since a matrix $\mathcal{P}$ has 12 components defined homogeneously up to a scale factor. In the following sections we derive view-based geometric entities that capture these 11 degrees of freedom in a form directly recoverable from matching image points measured in uncalibrated images. The recoverable quantities are called projective parameters, or epipolar parameters for two views. Section 2.6 outlines known methods for using projective parameters to obtain information about the Euclidean camera geometry.

2.2 Two Views

This section examines the projective relations between two views. We give a geometric derivation of the constraints on two matching image points in terms of two quantities, the epipoles and the epipolar collineation. We show how to express the constraint as a bilinear equation in the image point coordinates using the fundamental matrix. We show how to derive a minimally parameterized description of the epipoles and the epipolar collineation, also called the epipolar parameters. Finally, we show the explicit relationship between the epipolar parameters and the fundamental matrix.

2.2.1 Geometry of Two Views

Consider two camera views denoted by the indices j and k, with optical centers $\mathcal{C}_j$ and $\mathcal{C}_k$. Each optical center is projected into the other view's image plane at a distinguished point

called the *epipole*, denoted by $\mathbf{e}_{jk}$ and $\mathbf{e}_{kj}$, respectively, as shown in Figure 2-2. When a point $\mathbf{M}$ in the scene is seen in both views, it is projected along the two optical rays $< \mathcal{C}_j, \mathbf{M} >$ and $< \mathcal{C}_k, \mathbf{M} >$ to the two corresponding image points $\mathbf{m}_j$ and $\mathbf{m}_k$. The point $\mathbf{M}$ lies on both optical rays, and the triplet $< \mathcal{C}_j, \mathcal{C}_k, \mathbf{M} >$ defines a plane called the *epipolar plane*.

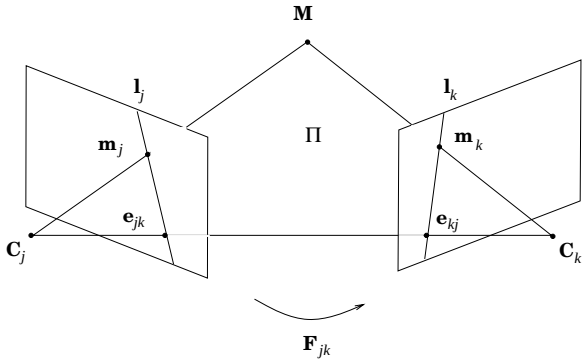


Figure 2-2: *The geometry of two views. The two optical rays* $< \mathbf{C}_j, \mathbf{M} >$ *and* $< \mathbf{C}_k, \mathbf{M} >$ *define the epipolar plane* $\mathbf{\Pi}$. *The projection of the epipolar plane into an image is an epipolar line.*

An important geometric construct for understanding the relationships between view pairs is the *epipolar line* through an image point $\mathbf{m}$. It is the projection of the epipolar plane containing the scene point $\mathbf{M}$ into an image plane. The epipolar line may also be thought of as the projection of one camera's optical ray into the other camera's image plane. Note that by construction, every epipolar line passes through the epipole in that image.

The one-parameter family of planes containing the line through $\mathcal{C}_j$ and $\mathcal{C}_k$ is called the pencil of epipolar planes. It induces a family of epipolar lines in each image called the *epipolar pencil*. Figure 2-3 depicts how the projection of each epipolar plane into an image is an epipolar line. The transformation that maps a line in the epipolar pencil of one image to its corresponding line in the other image is called the *epipolar transformation*. A pencil of lines in $\mathbb{P}^2$ is projectively equivalent to the projective line $\mathbb{P}^1$ [9]. Therefore, the epipolar transformation may be viewed as a collineation, an invertible transformation of $\mathbb{P}^1$ that is

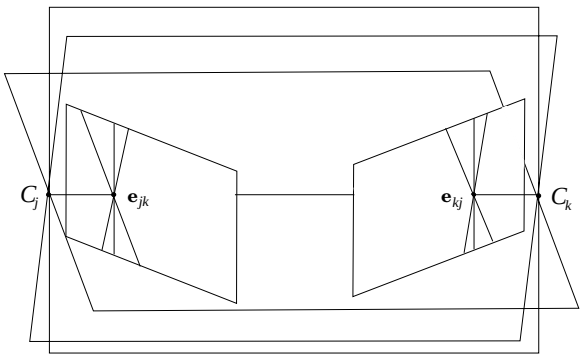


Figure 2-3: *The pencil of epipolar planes and the corresponding pencils of epipolar lines in two views. A pair of epipolar lines in views j and k are in correspondence through the epipolar collineation if they are images of the same epipolar plane.*

completely defined by three matching points, or in this case by three matching lines of the epipolar pencils [8]. We will also refer to it as the *epipolar collineation* between two views.

The epipoles and the epipolar collineation together form what is called the *epipolar geometry* of two views. They are jointly determined by the rotation between two views, the direction of translation between two views, and the intrinsic parameters of each view. They completely define the relative geometry of two uncalibrated views within the projective framework. To describe angles and distances, it is necessary to calibrate the views and move to the Euclidean framework. In the remainder of this work, we will refer to the minimum set of parameters that describe the epipoles and the epipolar collineation as the *epipolar parameters.*

2.2.2 Algebraic Constraints on Point Pairs

The relation between a pair of matching points in two images that is determined by the epipolar geometry of the two views may be captured as a multilinear relationship using the fundamental matrix. The fundamental matrix expresses the coplanarity constraint defined in Figure 1-4 as the transformation of a point $\mathbf{m}_j$ in one image to the epipolar line in the other image, the line through the epipole that contains the corresponding point $\mathbf{m}_k$. This

section derives the algebraic form describing this mapping, known as the *epipolar constraint*, in terms of the fundamental matrix, and shows its relation to the epipolar parameters.

Let two views of a scene, view j and view k, have camera projection matrices $\mathcal{P}_j \simeq [\mathbf{P}_j \quad \mathbf{p}_j]$ and $\mathcal{P}_k \simeq [\mathbf{P}_k \quad \mathbf{p}_k]$, respectively. Since the epipole is the image in one view of the other camera's optical center, we can express it by substituting the optical center in Equation 2.2 into the camera projection equation, Equation 2.1:

$$\mathbf{e}_{jk} \simeq \mathcal{P}_j \mathcal{C}_k \simeq -\mathbf{P}_j \mathbf{P}_k^{-1} \mathbf{p}_k + \mathbf{p}_j \tag{2.5}$$

$$\mathbf{e}_{kj} \simeq \mathcal{P}_k \mathcal{C}_j \simeq -\mathbf{P}_k \mathbf{P}_j^{-1} \mathbf{p}_j + \mathbf{p}_k. \tag{2.6}$$

Now let $\mathbf{m}_j \in \mathbb{P}^2$ and $\mathbf{m}_k \in \mathbb{P}^2$ be the images of a scene point $\mathbf{M} \in \mathbb{P}^3$ in each camera's image plane:

$$\mathbf{m}_j \simeq [\mathbf{P}_j \quad \mathbf{p}_j]\mathbf{M}$$

$$\mathbf{m}_k \simeq [\mathbf{P}_k \quad \mathbf{p}_k]\mathbf{M}.$$

We wish to obtain an expression relating these matching points via the epipolar transformation. This expression will both define the fundamental matrix and illustrate its relation to the epipoles and epipolar collineation.

We use the fact that in homogeneous coordinates, a projective line may be expressed as the cross-product of any two of its points. Let the notation $[.]_\times$ denote the matrix that performs the same linear operation as the cross-product.[2] Then the epipolar line containing $\mathbf{m}_j$ and $\mathbf{e}_{jk}$ is written as

$$\mathbf{l}_j \simeq \mathbf{e}_{jk} \times \mathbf{m}_j \simeq [\mathbf{e}_{jk}]_\times \mathbf{m}_j \tag{2.7}$$

$$\tag{2.8}$$

[2]The cross product of a vector $\mathbf{v} = (v_1, v_2, v_3)^T$ with another vector $\mathbf{u}$ may be written as a linear transformation $\mathbf{v} \times \mathbf{u} = [\mathbf{v}]_\times \mathbf{u}$, represented by the antisymmetric matrix $[\mathbf{v}]_\times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$.

and likewise the epipolar line through $\mathbf{m}_k$ and $\mathbf{e}_{kj}$ is

$$\mathbf{l}_k \simeq \mathbf{e}_{kj} \times \mathbf{m}_k \simeq [\mathbf{e}_{kj}]_\times \mathbf{m}_k. \tag{2.9}$$

To map $\mathbf{l}_j$ to $\mathbf{l}_k$, we use the fact that the 4×3 transpose $\mathcal{P}^T$ of a camera matrix performs the inverse projection of a line $\mathbf{l}$ in an image to an entire scene plane via the expression $\mathcal{P}^T\mathbf{l}$ [9]. The epipolar plane is the inverse projection of either epipolar line, so we can write it as

$$\mathbf{\Pi} \simeq \mathcal{P}_j^T \mathbf{l}_j.$$

$\mathbf{\Pi}$ is a 4-vector representing the equation of the plane (Section 2.1.1), so its dual in $\mathbb{P}^3$ is a projective point. Its projection in view k is the dual of the epipolar line,

$$\mathbf{l}_k \simeq \mathcal{P}_k \mathbf{\Pi} \simeq \mathcal{P}_k \mathcal{P}_j^T \mathbf{l}_j. \tag{2.10}$$

Writing the composition of the inverse and forward projections as the 3×3 matrix

$$\mathbf{H}_{jk} \simeq \mathcal{P}_k \mathcal{P}_j^T, \tag{2.11}$$

we obtain an algebraic form for the epipolar transformation that maps epipolar lines from view j to view k:

$$\mathbf{l}_k \simeq \mathbf{H}_{jk} \mathbf{l}_j. \tag{2.12}$$

The 3×3 matrix $\mathbf{H}_{jk}$ is not a unique representation of the epipolar transformation. Any matrix of the form $\mathbf{H}_{jk} + \mathbf{e}_{jk}\mathbf{a}^T$, where $\mathbf{a}$ is any 3-vector, also maps $\mathbf{l}_j$ to $\mathbf{l}_k$. The next section gives alternate representations of the epipolar transformation.

To obtain a constraint on the matching points $\mathbf{m}_j$ and $\mathbf{m}_k$ using the epipolar transformation, substitute the expressions of the epipolar lines from Equations 2.7 and 2.9 into

Equation 2.12 to obtain

$$\begin{aligned} [\mathbf{e}_{kj}]_\times \mathbf{m}_k &\simeq \mathbf{H}_{jk}[\mathbf{e}_{jk}]_\times \mathbf{m}_j \\ \mathbf{m}_k^T \mathbf{H}_{jk}[\mathbf{e}_{jk}]_\times \mathbf{m}_j &= 0. \end{aligned}$$

The *fundamental matrix* is defined as

$$\mathbf{F}_{jk} \simeq \mathbf{H}_{jk}[\mathbf{e}_{jk}]_\times, \tag{2.13}$$

and the *epipolar constraint* on matching points is written as

$$\mathbf{m}_k^T \mathbf{F}_{jk} \mathbf{m}_j = 0. \tag{2.14}$$

This equation is satisfied by two points in images j and k that are images of the same 3D scene point. The fundamental matrix thus maps a point from view j to its epipolar line in view k [17, 7].

By taking the transpose of Equation 2.14, we obtain an equivalent epipolar constraint with the roles of j and k swapped, $\mathbf{m}_j^T \mathbf{F}_{jk}^T \mathbf{m}_k = 0$, so we have the relation $\mathbf{F}_{kj} \simeq \mathbf{F}_{jk}^T$.

The fundamental matrix may also be expressed directly in terms of the internal parameter matrices and the Euclidean rotations and translations by expanding the epipole and epipolar transformations (Equations 2.5, 2.6, and 2.11) using the factored projection matrix in Equation 2.3. This form is presented in the discussion on projective-to-Euclidean mappings in Section 4.2.

Note that the 9 components of $\mathbf{F}_{jk}$ have only 7 degrees of freedom. The fact that the fundamental matrix is defined only up to a nonzero scale factor accounts for one less degree of freedom. In addition, it is clear from Equation 2.28 that the two epipoles belong to the right and left null spaces of the fundamental matrix, $\mathbf{F}_{jk}\mathbf{e}_{jk} = \mathbf{0}$ and $\mathbf{e}_{kj}^T \mathbf{F}_{jk} = (\mathbf{F}_{kj}\mathbf{e}_{kj})^T = 0$. Therefore, the fundamental matrix is nonsingular, so the constraint $\det(\mathbf{F}_{jk}) = 0$, a third-order constraint on the components of the fundamental matrix, accounts for the other missing degree of freedom.

The 3×3 fundamental matrix has only 7 degrees of freedom: its components must

satisfy two constraints. First, it is defined only up to a nonzero scale factor. Second, it is a singular matrix. It is clear from Equation 2.13 that $\mathbf{F}_{jk}\mathbf{e}_{jk} = \mathbf{0}$, so the epipole belongs to the null space of the fundamental matrix. The next section presents a minimal, 7-parameter representation of the epipolar geometry that corresponds to a decomposition of the fundamental matrix into its free parameters.

2.2.3 The Epipolar Parameters

This section shows how the epipolar geometry may be described by 7 free epipolar parameters, namely the 2 free parameters of each of the epipoles and the 3 free parameters of the epipolar collineation. We show how these parameters express the coplanarity constraint in terms of the epipolar transformation, and how they may be used to decompose the fundamental matrix into its minimally parameterized form [9].

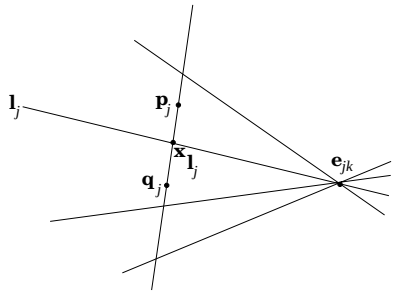


Figure 2-4: *The epipolar pencil through* $\mathbf{e}_{jk}$ *and its intersection* $\mathbf{x}_{\mathbf{l}_j}$ *with the transverse line* $< \mathbf{p}_j, \mathbf{q}_j >$.

Let h_{jk} denote the function mapping epipolar lines from view j to view k. When treating the epipolar lines as members of a pencil, we will express h_{jk} as a homogeneous 2×2 matrix, $\mathbf{H}_{jk_{2\times 2}}$ that operates on the epipolar pencil. To express the epipolar pencil as a one-parameter family, first define a transverse line in view j that does not contain the epipole $\mathbf{e}_{jk}$ by using two arbitrary points $\mathbf{p}_j$ and $\mathbf{q}_j$. Now identify each epipolar line in view j with its point of intersection $\mathbf{x}_{\mathbf{l}_j}$ with the transverse line (Figure 2-4). Analogously, in view k, choose two points $\mathbf{p}_k$ and $\mathbf{q}_k$ that form a line not containing the epipole $\mathbf{e}_{kj}$. An epipolar line in view k is represented by its intersection $\mathbf{x}_{\mathbf{l}_k}$ with the transverse line.

The point $\mathbf{x}_{\mathbf{l}_j}$ (resp. $\mathbf{x}_{\mathbf{l}_k}$) may now be expressed as a linear combination of the basis points $\mathbf{p}_j$ and $\mathbf{q}_j$ (resp. $\mathbf{p}_k$ and $\mathbf{q}_k$). An epipolar line $\mathbf{l}_j$ is thus defined by the coefficients

α_j and β_j of its intersection point $\mathbf{x}_{\mathbf{l}_j}$:

$$\begin{aligned} \mathbf{x}_{\mathbf{l}_j} &\simeq \mathbf{l}_j \times (\mathbf{p}_j \times \mathbf{q}_j) \\ &\simeq (\mathbf{l}_j^T \mathbf{q}_j)\mathbf{p}_j - (\mathbf{l}_j^T \mathbf{p}_j)\mathbf{q}_j, \end{aligned}$$

so

$$\mathbf{x}_{\mathbf{l}_j} \simeq \alpha_j \mathbf{p}_j + \beta_j \mathbf{q}_j, \tag{2.15}$$

where

$$\alpha_j = \mathbf{l}_j^T \mathbf{q}_j \tag{2.16}$$

$$\beta_j = -\mathbf{l}_j^T \mathbf{p}_j. \tag{2.17}$$

These coefficients may be written as the homogeneous coordinates $\mathbf{x}_{\mathbf{l}_j} \simeq [\alpha_j, \beta_j]^T$ of a point on a projective line. Similarly, in view k, the coordinates of the point $\mathbf{x}_{\mathbf{l}_k}$ representing $\mathbf{l}_k$ are

$$\alpha_k = \mathbf{l}_k^T \mathbf{q}_k \tag{2.18}$$

$$\beta_k = -\mathbf{l}_k^T \mathbf{p}_k, \tag{2.19}$$

and

$$\mathbf{x}_{\mathbf{l}_k} \simeq \alpha_k \mathbf{p}_k + \beta_k \mathbf{q}_k. \tag{2.20}$$

We can now write the epipolar collineation as a homogeneous matrix

$$\mathbf{H}_{jk_{2\times 2}} \simeq \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{2.21}$$

that maps corresponding epipolar lines to each other by transforming their coefficients as

follows:

$$\mathbf{H}_{jk_{2\times 2}} \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \simeq \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \simeq \begin{bmatrix} a\alpha_j + b\beta_j \\ c\alpha_j + d\beta_j \end{bmatrix} \simeq \begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix}.$$

Expressed as an equality, we have

$$\frac{\alpha_k}{\beta_k} = \frac{a\alpha_j + b\beta_j}{c\alpha_j + d\beta_j}, \tag{2.22}$$

so

$$a(\alpha_j\beta_k) + b(\beta_j\beta_k) - c(\alpha_j\alpha_k) - d(\beta_j\alpha_k) = 0. \tag{2.23}$$

This form is the minimally parameterized representation of how the epipolar collineation maps epipolar lines. Each epipolar line is represented as a member of it respective epipolar pencil by two homogeneous coordinates, $[\alpha_j, \beta_j]^T$ (resp. $[\alpha_k, \beta_k]^T$), and the epipolar transformation is represented by its four homogeneous components. The practical issues that arise from this minimal parameterization of the epipolar geometry are discussed in Section 5.4.

Now that we have a parameterized representation of the epipolar pencil, we can express the epipolar constraint on matching image points in terms of the minimal set of epipolar parameters. Again, let $\mathbf{m}_j$ and $\mathbf{m}_k$ be matching points in views j and k lying on their epipolar lines, $\mathbf{l}_j$ and $\mathbf{l}_k$, respectively. Now substitute the expressions for the epipolar lines from Equations 2.7 and 2.9 into Equations 2.16 to 2.19 to obtain the following expressions for $\mathbf{x}_{\mathbf{l}_j} \simeq [\alpha_j, \beta_j]^T$ and $\mathbf{x}_{\mathbf{l}_k} \simeq [\alpha_k, \beta_k]^T$ in terms of the point matches,

$$\alpha_j = (\mathbf{e}_{jk} \times \mathbf{m}_j)^T \mathbf{q}_j = \mathbf{q}_j^T [\mathbf{e}_{jk}]_\times \mathbf{m}_j \tag{2.24}$$

$$\beta_j = -(\mathbf{e}_{jk} \times \mathbf{m}_j)^T \mathbf{p}_j = -\mathbf{p}_j^T [\mathbf{e}_{jk}]_\times \mathbf{m}_j \tag{2.25}$$

$$\alpha_k = (\mathbf{e}_{kj} \times \mathbf{m}_k)^T \mathbf{q}_k = -\mathbf{m}_k^T [\mathbf{e}_{kj}]_\times \mathbf{q}_k \tag{2.26}$$

$$\beta_k = -(\mathbf{e}_{kj} \times \mathbf{m}_k)^T \mathbf{p}_k = \mathbf{m}_k^T [\mathbf{e}_{kj}]_\times \mathbf{p}_k. \tag{2.27}$$

These α's and β's represent the epipolar lines as members of their epipolar pencil.

Substituting them into Equation 2.22, we get the expression,

$$\begin{aligned}(\mathbf{m}_k^T[\mathbf{e}_{kj}]_\times\mathbf{p}_k)a(\mathbf{q}_j^T[\mathbf{e}_{jk}]_\times\mathbf{m}_j) + (\mathbf{m}_k^T[\mathbf{e}_{kj}]_\times\mathbf{p}_k)b(-\mathbf{p}_j^T[\mathbf{e}_{jk}]_\times\mathbf{m}_j)\\ -(-\mathbf{m}_k^T[\mathbf{e}_{kj}]_\times\mathbf{q}_k)c(\mathbf{q}_j^T[\mathbf{e}_{jk}]_\times\mathbf{m}_j) - (-\mathbf{m}_k^T[\mathbf{e}_{kj}]_\times\mathbf{q}_k)d(-\mathbf{p}_j^T[\mathbf{e}_{jk}]_\times\mathbf{m}_j) \quad = \quad 0.\end{aligned}$$

By factoring out the points $\mathbf{m}_j$ and $\mathbf{m}_k$ and the epipoles $\mathbf{e}_{jk}$ and $\mathbf{e}_{kj}$, we simplify this equality to the standard form of the epipolar constraint on matching pairs of image points, $\mathbf{m}_k^T\mathbf{F}_{jk}\mathbf{m}_j = 0$, where now the fundamental matrix is decomposed as,

$$\mathbf{F}_{jk} \quad \simeq \quad [\mathbf{e}_{kj}]_\times[\mathbf{p}_k \;\; \mathbf{q}_k]\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} \mathbf{q}_j^T \\ -\mathbf{p}_j^T \end{bmatrix}[\mathbf{e}_{jk}]_\times. \tag{2.28}$$

This expression, due to [9], shows how the fundamental matrix depends precisely on 7 epipolar parameters that describe the epipolar geometry: 2 parameters for each epipole, and 3 parameters for the epipolar collineation (Figure 2-5).

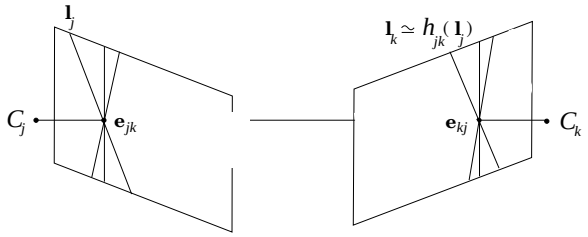


Figure 2-5: *The 7 projective parameters defining the epipolar geometry, 2 for each epipole, $\mathbf{e}_{jk}$ and $\mathbf{e}_{kj}$ and 3 for the epipolar collineation h_{jk}. Two lines, $\mathbf{l}_j$ and $\mathbf{l}_k$, are in correspondence through h_{jk} if they lie in the same epipolar plane (see Figure 2-3).*

This decomposition is not unique. There are infinitely many 7-parameter decompositions, all corresponding to the same fundamental matrix, depending on how the epipoles' homogeneous coordinates are normalized and on the basis chosen for the epipolar pencils. The next section describes a particular set of instantiations of this parameterization, and Section 5.4 explores the geometric implications of the various parameterization choices.

2.2.4 Canonical Parameterizations

Equation 2.28 expresses the fundamental matrix in terms of a collineation $\mathbf{H}_{jk_{2\times 2}}$ that depends on an arbitrary choice of basis points, $\mathbf{p}_j$, $\mathbf{q}_j$, $\mathbf{p}_k$, and $\mathbf{q}_k$. Since there are infinitely many choices of basis points, there are infinitely many possible parameterizations of the epipolar geometry.

Csurka, *et al.* give a finite set of parameterizations of the fundamental matrix, each corresponding to a different normalization component for the epipoles and the epipolar collineation [5]. We show here how these parameterizations are equivalent to restricting the representations of the epipolar pencil by limiting the basis points to a set of canonical values,

$$\boldsymbol{\epsilon}_1 \simeq [1,0,0]^T \tag{2.29}$$

$$\boldsymbol{\epsilon}_2 \simeq [0,1,0]^T \tag{2.30}$$

$$\boldsymbol{\epsilon}_3 \simeq [0,0,1]^T \tag{2.31}$$

The resulting parameterizations are useful in practice because they permit the epipolar collineation to be immediately observable from a given fundamental matrix.

First choose the indices to be used for normalizing the two epipoles: $\zeta_j = [j_1, j_2, j_3]^T$ and $\zeta_k = [k_1, k_2, k_3]^T$. Let j_3 be the index of $\mathbf{e}_{jk}$ chosen for normalization, i.e., $\mathbf{e}_{jk}(j_3) = 1$, and let j_1 and j_2 denote the other two indices ($j_1 < j_2$). Now assign the points $\mathbf{p}_j$ and $\mathbf{q}_j$ using the canonical basis elements $\boldsymbol{\epsilon}_{j_1}$ and $\boldsymbol{\epsilon}_{j_2}$:

$$\mathbf{p}_j \simeq (-1)^{(j_3)}\boldsymbol{\epsilon}_{j_1} \tag{2.32}$$

$$\mathbf{q}_j \simeq (-1)^{(j_3)}\boldsymbol{\epsilon}_{j_2}. \tag{2.33}$$

For view k, let k_3 be the index of $\mathbf{e}_{kj}$ for normalization, so $\mathbf{e}_{kj}(j_3) = 1$, and k_1 and k_2 be the other two indices ($k_1 < k_2$). Let

$$\mathbf{p}_k \simeq -(-1)^{(k_3)}\boldsymbol{\epsilon}_{k_2} \tag{2.34}$$

$$\mathbf{q}_k \simeq (-1)^{(k_3)}\boldsymbol{\epsilon}_{k_1}. \tag{2.35}$$

Now by substituting these values into the parameterization in Equation 2.28, we can verify that the components of the collineation are components of the fundamental matrix:

$$a = \mathbf{F}_{jk}(k_1, j_1) \tag{2.36}$$

$$b = \mathbf{F}_{jk}(k_1, j_2) \tag{2.37}$$

$$c = \mathbf{F}_{jk}(k_2, j_1) \tag{2.38}$$

$$d = \mathbf{F}_{jk}(k_2, j_2). \tag{2.39}$$

We will refer to the parameterizations that use canonical basis points for the epipolar pencil as *canonical parameterizations*. The next section instantiates a particular canonical parameterization that is used in later chapters.

2.2.5 The Canonical Parameterization

As an example of a canonical parameterization, choose the epipole normalization indices to be $j_3 = 3$ and $k_3 = 3$, so $\mathbf{e}_{jk} = (e_1, e_2, 1)^T$ and $\mathbf{e}_{kj} = (e'_1, e'_2, 1)^T$. Then $(j_1, j_2) = (1, 2)$ and $(k_1, k_2) = (1, 2)$, so

$$\begin{aligned}
\mathbf{p}_j &\simeq (-1)^3\boldsymbol{\epsilon}_1 &&= (-1, 0, 0)^T \\
\mathbf{q}_j &\simeq (-1)^3\boldsymbol{\epsilon}_2 &&= (0, -1, 0)^T \\
\mathbf{p}_k &\simeq -(-1)^3\boldsymbol{\epsilon}_2 &&= (0, 1, 0)^T \\
\mathbf{q}_k &\simeq (-1)^3\boldsymbol{\epsilon}_1 &&= (-1, 0, 0)^T.
\end{aligned}$$

Note that these are points at infinity, so the transverse line is in fact the line at infinity, parameterized by the points at infinity in the horizontal and vertical directions. Substituting these values into Equation 2.28, we obtain a formula for $\mathbf{F}_{jk}$ in which the parameters of the collineation, $\{a, b, c, d\}$, are immediately observable:

$$\mathbf{F}_{jk} \simeq [\mathbf{e}_{kj}]_\times \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} [\mathbf{e}_{jk}]_\times$$

$$
\begin{aligned}
&\simeq [\mathbf{e}_{kj}]_\times \begin{bmatrix} -d & c & 0 \\ b & -a & 0 \\ 0 & 0 & 0 \end{bmatrix} [\mathbf{e}_{jk}]_\times \\
&\simeq \begin{bmatrix} a & b & -(ae_1 + be_2) \\ c & d & -(ce_1 + de_2) \\ -(ae'_1 + ce'_2) & -(be'_1 + de'_2) & (ae_1 + be_2)e'_1 + (ce_1 + de_2)e'_2 \end{bmatrix}.
\end{aligned}
$$

Since the collineation is defined only up to scale, we represent it using only 3 parameters by defining the indices $\zeta_c = [c_1, c_2, c_3, c_4]^T$ into $\mathbf{h}_{jk}$, the unrolled vector form of $\mathbf{H}_{jk_{2\times 2}}$, and then normalizing $\mathbf{H}_{jk_{2\times 2}}$ by the c_4^{th} component of $\mathbf{h}_{jk}$, so $\mathbf{h}_{jk}(c_4) = 1$.

This particular choice of normalization indices and epipolar collineation bases gives us the minimal parameterization used in [5] and [53]. We will often refer to it as the canonical parameterization.

2.2.6 Summary

This section has given a description of the epipolar geometry in terms of 7 epipolar parameters,

$$
\mathbf{p}^7_{jk} = \{\mathbf{e}_{jk}(j_2), \mathbf{e}_{jk}(j_3), \mathbf{e}_{kj}(k_2), \mathbf{e}_{kj}(k_3), \mathbf{h}_{jk}(c_1), \mathbf{h}_{jk}(c_2), \mathbf{h}_{jk}(c_3)\}. \tag{2.40}
$$

that depend on a set of index vectors ζ_j, ζ_k, and ζ_c. In addition, we have established procedures for decomposing a fundamental matrix $\mathbf{F}_{jk}$ into 7 minimal parameters, and vice-versa, for constructing a fundamental matrix from a parameter set $\mathbf{p}^7$ and three index vectors.

To make sure the parameterization is well-defined, we require that j_3 and k_3 be chosen so that $\mathbf{e}_{jk}(j_3) \neq 0$ and $\mathbf{e}_{kj}(k_3) \neq 0$. Likewise, c_4 must be chosen so that $\mathbf{h}_{jk}(c_4) \neq 0$. There are three ways of choosing each epipole's normalization component, $j_1 = 1, 2, 3$ and $k_1 = 1, 2, 3$, and four ways of choosing the collineation's normalization component, $c_4 = 1, 2, 3, 4$, for a total of $3 \times 3 \times 4 = 36$ possible canonical parameterizations.

While a collineation may in general be represented by any two basis points, the finitely

many choices given by the canonical basis points lead to analytical expressions for extracting the collineation components directly from a fundamental matrix by observation.

Since there are multiple parameterizations, the mapping from $\mathbf{F}_{jk}$ to $\mathbf{p}^7_{jk}$ is one-to-many. Although in theory, all parameter sets give the same error on matching data points, the manifolds of these parameter sets are not identical, so the optimization of an objective function over the parameters may have distinctly different behavior depending on the chosen representation. Statistical methods have been studied for choosing the most numerically favorable normalization indices at any given iteration. In [5], the normalization indices are chosen to maximize the rank of the Jacobian of the function from $\mathbf{F}_{jk}$ to $\mathbf{p}^7_{jk}$. Chapter 5 presents a more general parameterization that allows an arbitrary choice basis for the epipolar collineation.

Various forms of the epipole-collineation parameterization of view pairs have been used in fundamental matrix estimation and in projective bundle adjustment [2]. The new minimal model of three views presented in Chapter 3 will be constructed from the dependencies that exist between collections of these epipolar parameters for different view pairs.

2.3 Three Views

This section presents existing models for the geometry of three uncalibrated views. First, we construct the trifocal relationships geometrically. Then we show how point triplets are constrained by the trifocal geometry, first using the trifocal tensor, and then directly in terms of the pairwise epipolar parameters, when the optical centers and scene point are in general position. This framework sets the stage for the minimally parameterized trifocal model introduced in Chapter 3.

2.3.1 Geometry of Three Views

Given three views i, j, and k from cameras whose optical centers are in general position, what constraints exist on the triplet of matching image points, $\mathbf{m}_i$, $\mathbf{m}_j$, and $\mathbf{m}_k$ corresponding to a single 3D point in terms of the camera configuration? Figure 2-6 illustrates that if three image points are in correspondence across the three views, then their optical rays

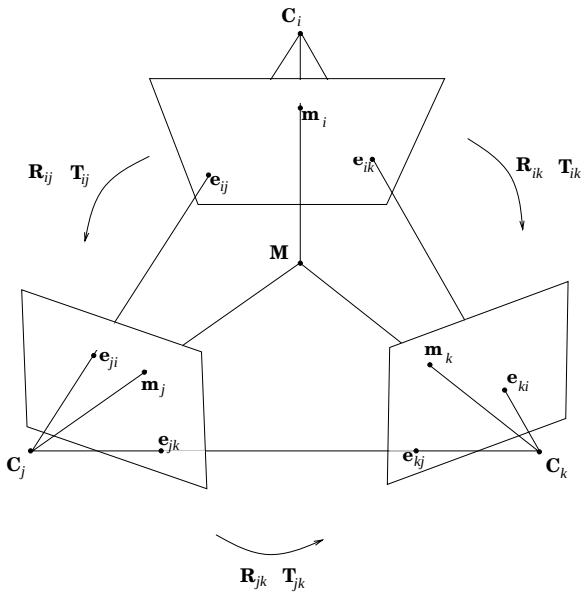


Figure 2-6: *The geometry of three views.*

must intersect at a single point $\mathbf{M}$ in the scene. The plane containing the three optical centers $\mathcal{C}_i$, $\mathcal{C}_j$, and $\mathcal{C}_k$ is called the *trifocal plane*, and its intersections with the three image planes are called the *trifocal lines*, $\mathbf{t}_i$, $\mathbf{t}_j$, and $\mathbf{t}_k$, as shown in Figure 2-7

The following sections present two different algebraic representations of the trifocal geometry and the constraints it places on matching points in three views. First we present the trifocal tensor and its associated trilinear constraints on point triplets. The trifocal tensor is a convenient algebraic tool for modeling view triplets, but we will see that its linear solution is not sufficiently constrained, *i.e.*, there is a lower-dimensional nonlinear manifold embedded in its linear solution space on which the true solution must lie.

As an alternative to the trifocal tensor, the trifocal geometry may be modeled as three instances of the epipolar geometry when the camera centers lie in general position. It will be important to note the dependencies among the three fundamental matrices, since they will be used in different forms in Chapter 3 to model three views with a minimal collection of epipolar parameters.

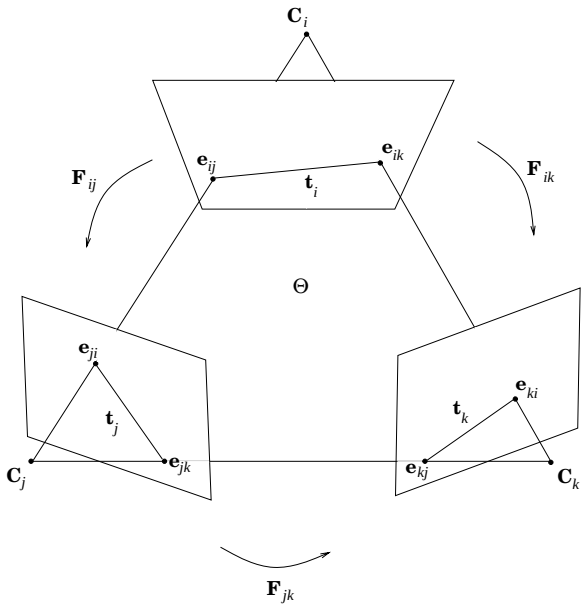


Figure 2-7: *The trifocal plane* $\mathbf{\Theta}$ *is defined by the three optical centers,* $< \mathcal{C}_i, \mathcal{C}_j, \mathcal{C}_k >$*. Each trifocal line is the projection of the trifocal plane into the image plane. It is also an epipolar line for each of the two other views.*

2.3.2 Algebraic Constraints on Point Triplets

The previous sections showed how two images of a point are constrained by the 7 epipolar parameters, and how they define the epipolar geometry. Whereas the fundamental matrix that corresponds to these epipolar parameters provides a bilinear expression for the constraints that the epipolar geometry imposes on matching point pairs, the trifocal tensor provides an analogous constraint on matching point triplets.

The trifocal tensor trilinearly relates triplets of both points and lines, where a point in one image is considered to match any line in another image that contains its matching point. The trifocal tensor is a $3 \times 3 \times 3$ array that captures the constraints on point triplets or line triplets or combinations of points and lines. Its 27 components are coefficients of a set of trilinear expressions in the projective coordinates of image points or lines. These coefficients are not independent: they must satisfy constraints that constrain the tensor to lie on on an 18-dimensional manifold of admissible values [12]. However, the tensor is an attractive model for view triplets since its elements may be recovered by linear methods from matching image triplets.

The Line Triplet Constraint

Let us first describe the trifocal tensor as a constraint on line matches, since its immediate corollaries provide constraints on matching point triplets [18, 13]. Let $\mathcal{P}_i$, $\mathcal{P}_j$, and $\mathcal{P}_k$ be the three projection matrices of three cameras with distinct optical centers. Without loss of generality, choose a coordinate system such that, $\mathcal{P}_i \simeq [\mathbf{I}\ \mathbf{0}]$, and partition the other two projection matrices as $\mathcal{P}_j \simeq [\mathbf{P}_j\ \ \mathbf{p}_j]$ and $\mathcal{P}_k \simeq [\mathbf{P}_k\ \ \mathbf{p}_k]$.

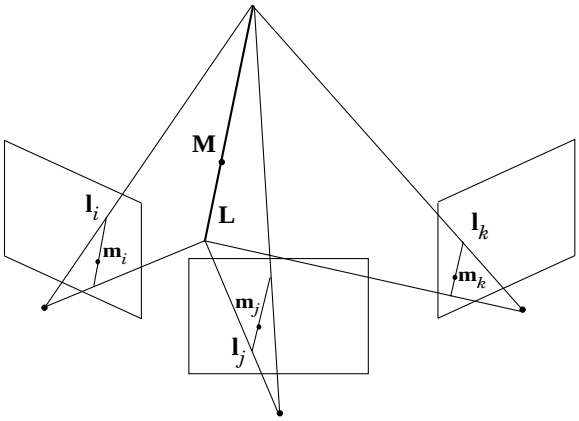


Figure 2-8: *The geometry of three views of a line.*

Let $\mathbf{l}_i$, $\mathbf{l}_j$, and $\mathbf{l}_k$ be corresponding lines in each of the three camera images, *i.e.,* images of a single line $\mathbf{L}$ in space (Figure 2-8). Note that there is an entire plane in $\mathbb{P}^3$ whose image is $\mathbf{l}_i$, and the same is true for the other two cameras. These planes may be represented as a back-projection of the image lines and written purely in terms of the camera matrices and line coefficients. The requirement that these three planes intersect in a unique line may thus be stated in terms of the image lines and the camera matrices. The resulting relation between image lines may be rearranged as a mapping $\mathcal{T}_i^{jk}$ that, given a pair of lines $\mathbf{l}_j$ and $\mathbf{l}_k$ from views j and k, returns the corresponding line $\mathbf{l}_i$ in view i:

$$\mathbf{l}_i \simeq \mathcal{T}_i^{jk}(\mathbf{l}_j, \mathbf{l}_k) \simeq \begin{bmatrix} \mathbf{l}_j{}^T \mathbf{G}_i^1 \mathbf{l}_k \\ \mathbf{l}_j{}^T \mathbf{G}_i^2 \mathbf{l}_k \\ \mathbf{l}_j{}^T \mathbf{G}_i^3 \mathbf{l}_k \end{bmatrix} \tag{2.41}$$

The notation $\mathcal{T}_i^{jk}$ often refers to the $3 \times 3 \times 3$ trifocal tensor whose three slices are the

three matrices $\mathbf{G}^1$, $\mathbf{G}^2$, and $\mathbf{G}^3$. The 27 components of the tensor slices are the coefficients of products of the line components. The three matrices were originally derived in the Euclidean framework as constraints on point triplets [38], and much later reformulated as a single trilinear tensor [37].

Taking the cross product of each side of Equation 2.41 with $\mathbf{l}_i$ gives a set of three multilinear equations in the line components which can be rearranged into a homogeneous system of linear equations in the 27 elements of the tensor $\mathcal{T}_i^{jk}$:

$$\mathbf{l}_i \times \begin{bmatrix} \mathbf{l}_j{}^T \mathbf{G}_i^1 \mathbf{l}_k \\ \mathbf{l}_j{}^T \mathbf{G}_i^2 \mathbf{l}_k \\ \mathbf{l}_j{}^T \mathbf{G}_i^3 \mathbf{l}_k \end{bmatrix} = \mathbf{0}.$$

The Point Triplet Constraint

From the trilinear constraints on lines we can immediately move to a set of trilinear constraints on points, by first deriving the constraints on a point-line-line match, *i.e.*, a point in one image and a line in each of the other two images that contains the corresponding image point. The point-line-line form directly follows from the fact that a point $\mathbf{m}$ on the line $\mathbf{l}$ must satisfy $\mathbf{m}^T\mathbf{l} = 0$, so if $\mathbf{M}$ is a point on the line $\mathbf{L}$ of the scene, and its image in camera i is the point $\mathbf{m}_i$ lying on $\mathbf{l}_i$, then

$$\mathbf{m}_i^T \mathbf{l}_i = \mathbf{m}_i^T \begin{bmatrix} \mathbf{l}_j^T \mathbf{G}_i^1 \mathbf{l}_k \\ \mathbf{l}_j^T \mathbf{G}_i^2 \mathbf{l}_k \\ \mathbf{l}_j^T \mathbf{G}_i^3 \mathbf{l}_k \end{bmatrix} = 0. \tag{2.42}$$

This equation is trilinear in the image coordinates of the lines $\mathbf{l}_j$ and $\mathbf{l}_k$ and the point $\mathbf{m}_i$.

To derive the relationship between the components of three corresponding image points, $\mathbf{m}_i$, $\mathbf{m}_j$, and $\mathbf{m}_k$, use the three basis vectors, $\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$, and $\boldsymbol{\epsilon}_3$, defined in Section 2.2.4. We can construct lines containing the matching points by defining $\mathbf{l}_j \simeq \boldsymbol{\epsilon}_m \times \mathbf{m}_j$ and $\mathbf{l}_k \simeq \boldsymbol{\epsilon}_n \times \mathbf{m}_k$, for any choice of $m, n \in \{1, 2, 3\}$. Any such pair of lines may be back-projected into the scene to two planes that intersect to form a unique line $\mathbf{L}$. The image of $\mathbf{L}$ in view i must contain the point $\mathbf{m}_i$, so we have a point-line-line match. Substituting $\mathbf{l}_j$ and $\mathbf{l}_k$ into

Equation 2.42 gives us a set of 9 trilinear equations (for all choices of m and n) in the elements of the three corresponding image points:

$$\mathbf{m}_i^T \begin{bmatrix} (\boldsymbol{\epsilon}_m \times \mathbf{m}_j)^T \mathbf{G}_i^1 (\boldsymbol{\epsilon}_n \times \mathbf{m}_k) \\ (\boldsymbol{\epsilon}_m \times \mathbf{m}_j)^T \mathbf{G}_i^2 (\boldsymbol{\epsilon}_n \times \mathbf{m}_k) \\ (\boldsymbol{\epsilon}_m \times \mathbf{m}_j)^T \mathbf{G}_i^3 (\boldsymbol{\epsilon}_n \times \mathbf{m}_k) \end{bmatrix} = 0. \tag{2.43}$$

Of these equations, four are linearly independent and are known as the *trilinearities* given by Shashua in [37].

Thus, the trifocal tensor is a linear model constraining point triplets. As mentioned earlier, its 27 components are algebraically related through their dependence on the camera projection matrices, so they must satisfy constraints that restrict the tensor to only 18 degrees of freedom. Enforcing these constraints has proved challenging and remains an active area of research [13].

2.3.3 Trifocally Constrained Fundamental Matrices

In this section, we ask two questions. First, under what conditions do the three fundamental matrices between view pairs impose the same constraints on matching points as the trifocal tensor? And secondly, under what conditions do the three sets of epipolar parameters contain the same information about the geometry of the cameras as does the trifocal tensor?

First note that the three fundamental matrices in a triplet are interdependent. The dependencies may be stated algebraically by treating the image locations of one camera center in each of the two other views as a matching point pair that must satisfy the epipolar constraint. The entries in Table 2.1 state all three permutations of this relationship. Chapter 3 presents alternate formulations of these dependencies for the purpose of deriving minimally parameterized trifocal models. These forms, expressed in terms of the trifocal lines, provide geometric methods for constraining the set of all epipolar parameters among the three pairs in a view triplet.

When the three optical centers are in general position, the three epipolar constraints completely encapsulate the relations among matching point triplets. A point in one view must lie on two different epipolar lines, one for each of the other two views. The intersection

$\mathbf{e}_{ij}^T \mathbf{F}_{ki} \mathbf{e}_{kj} = 0$	$\mathbf{e}_{ik}^T \mathbf{F}_{ji} \mathbf{e}_{jk} = 0$
$\mathbf{e}_{jk}^T \mathbf{F}_{ij} \mathbf{e}_{ik} = 0$	$\mathbf{e}_{ji}^T \mathbf{F}_{kj} \mathbf{e}_{ki} = 0$
$\mathbf{e}_{ki}^T \mathbf{F}_{jk} \mathbf{e}_{ji} = 0$	$\mathbf{e}_{kj}^T \mathbf{F}_{ik} \mathbf{e}_{ij} = 0$

Table 2.1: *The dependencies among three fundamental matrices in a view triplet.*

of the two epipolar lines completely determines the point's image position (Figure 2-9). This construction is identical to the point transfer property of the trifocal tensor, so the tensor offers no additional constraints on the three image points. The collection of epipolar parameters is therefore sufficient for modeling the trifocal geometry of three cameras in general position.

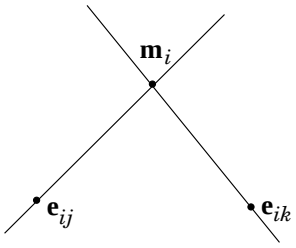


Figure 2-9: *Two epipolar lines corresponding to two matched points in other images intersect at the third matched point.*

2.3.4 Special Cases

The three epipolar geometries fail to constrain images of scene points that lie in the trifocal plane. In this case, the two epipolar lines for a given point correspondence coincide since they are identical to the trifocal line. Fortunately, it is unlikely in practice that all points viewed in a scene happen to lie on the trifocal plane.

When the three optical centers are collinear, the trifocal tensor places stronger constraints on point correspondence than the epipolar geometry does. For a collinear triplet of cameras, the three epipolar planes for any scene point are identical, so a point's epipolar line with respect to one view coincides with the epipolar line with respect to the other view. Figure 2-10 depicts this scenario: point transfer from two views to a third is no longer possible using only the epipolar relations since there are no longer two epipolar lines intersecting

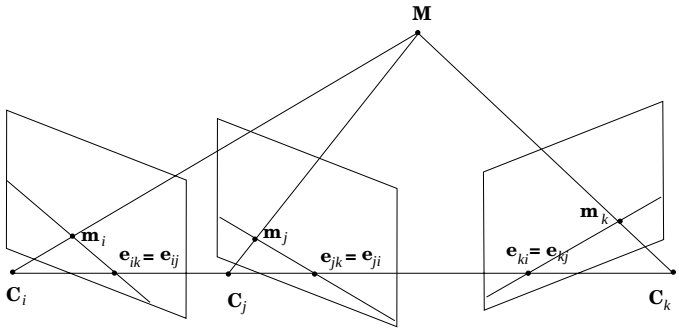


Figure 2-10: *Three views with collinear optical centers.*

at a unique point in the third view. The trifocal tensor, however, does constrain all three points in this case.

2.4 N Views

The generalization of the bifocal and trifocal constraints to N-focal projective constraints on image points across arbitrarily many views has been mathematically derived but is not commonly used in practice [6, 22, 45]. A simple algebraic formulation of these constraints begins with the N projection equations (Equation 2.1) that describe N views of a scene point:

$$\mathbf{m}_i \quad = \quad \lambda_i \mathcal{P}_i \mathbf{M}, \quad i = 1, 2, ...N.$$

When taken simultaneously, these may be rearranged into various systems of linear equations. Early work in multiple-view structure from motion for orthographic cameras rearranged this system into a factorizable matrix of camera properties and scene structure [41]. To obtain constraints on image points in terms of camera properties, the equations are

rearranged to form the homogeneous system of equations,

$$\begin{bmatrix} \mathcal{P}_1 & \mathbf{m}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathcal{P}_2 & \mathbf{0} & \mathbf{m}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & & & \vdots \\ \mathcal{P}_N & \mathbf{0} & \dots & \mathbf{0} & \mathbf{m}_N \end{bmatrix} \begin{bmatrix} \mathbf{M} \\ -\lambda_1 \\ -\lambda_2 \\ \vdots \\ -\lambda_N \end{bmatrix} = \mathbf{0}.$$

The minors of this rank-deficient matrix are multilinear constraints on matching image points, and the 3^N coefficients of these forms are the elements of a multilinear tensor, the N-view analog of the 9-dimensional fundamental matrix and the 27-dimensional trifocal tensor [45]. It turns out that the set of equations given by this tensor offer no additional additional constraints on matching image points than does the full set of trilinearities derived from the trifocal tensor for all possible triplets in the collection [21]. Furthermore, the suggested methods for extracting camera geometry from the multilinear tensor involve first determining the epipoles and fundamental matrices [46]. For these reasons, it has not been necessary to use the multilinear tensor in practice as a tool for estimating camera geometry.

2.5 Degrees of Freedom

Until now we have avoided the important concept of a "minimal" parameterization, i.e. a representation of the camera configuration that permits only the geometrically plausible degrees of freedom. For two uncalibrated views, we saw that there were 7 degrees of freedom. For N views, there are $11N - 15$ degrees of freedom: each camera may be described by the elements of its 3×4 projection matrix minus a single scale factor, providing 11 parameters, and the entire configuration is defined up to a 3D projective transformation, which has $4x4$ components defined up to scale, and thus 15 degrees of freedom. Alternatively, we can think of each camera as being defined in terms of its 3D position (3 parameters), its 3D rotation (3 parameters), and its intrinsic camera parameters (5 parameters) for a total of 11 parameters for each additional view.

Since there are only $11N-15$ degrees of freedom in a view collection, and 3^N components

in a multilinear tensor, the tensor components are clearly algebraically dependent. The linear solution to the unknown tensor components given a set of point matches across N views will not be guaranteed to lie in the true space of admissible descriptions of the configuration. There are $3^N-(11N-15)$ constraints relating the tensor elements which must somehow be enforced if a solution is to correspond to a projective camera configuration.

For 2 views, there are established methods for projecting the 9 homogeneous components of a fundamental matrix onto a 7-dimensional space of admissible solutions [19, 53]. For 3 views, there has been work on forcing the 27 redundant coefficients of the trifocal tensor to lie in the 18-dimensional space of admissible solutions by defining and enforcing 9 independent nonlinear constraints among the tensor coefficients [13]. Other methods minimally parameterize the trifocal tensor in terms of a set of actual image data points that are thought to be well-matched [43]. For 4 or more views, the exponentially many constraints on multilinear tensors have been derived algebraically, but they are redundant and nonlinear, and methods for enforcing them are not yet well-established [47].

2.6 Projective Structure and Motion

The geometry of multiple views has been modeled as a collection of epipolar relations by several others. Laveau, *et al.* use collections of fundamental matrices to implicitly represent 3D surfaces via novel view prediction [24]. Rothwell, *et al.* demonstrate how a projective 3D camera representation may be constructed from two uncalibrated views by finding the family of projection matrices that agree with an estimated fundamental matrix [34]. More generally, Luong, *et al.*, describe how this family may be normalized to define a minimally parameterized set of projection matrices in the projective framework [28].

The relations between multiple views have also been modeled as collections of trifocal tensors [16]. However, the resulting collection of three-view models is not globally consistent, so additional steps must be taken to minimize the disagreement between different camera pairs.

The factorization approach to projective structure and motion recovery is a generalization of the image factorization work of [41]. Sturm and Triggs [40] present a projective

extension of this work, along with an N-view model of multiple views called the "joint image." This method estimates the epipolar geometries for an entire collection along with the depths of all the observed points, and requires a set of measured image points that is visible in all views.

Bundle adjustment is a procedure for optimizing the 3D rotations and translations of many cameras while simultaneously estimating the 3D positions of scene points viewed in multiple images [4]. The bundle being adjusted for a given set of image matches is the set of optical rays originating at the current estimates of camera optical centers, passing through the image points, and ideally intersecting at a single 3D scene point. The algorithm alternates between computing the positions of 3D scene points given the current camera poses, and minimizing the distances of the projected image points to the measured point matches by adjusting the 3D camera rotations and translations. The procedure must be initialized with reasonably accurate Euclidean parameters if it is to converge to the global optimum.

Bundle adjustment may be performed in the projective framework as well as in the Euclidean framework. Camera centers, scene points, optical rays, image planes, and image points are defined projectively, and optical rays are now adjusted to intersect at a unique 3D projective point. Bartoli, *et al.* minimally parameterize the camera projection matrices in terms of the epipoles and epipolar transformations [2]. The 3D solutions to camera and scene positions are defined only up to an arbitrary projective transform. For a recent review of these methods, see [48].

2.7 Summary

This projective geometric framework for modeling multiple views is the basis for the model presented in the next chapter. Starting from the minimal representation of two views given here, we will derive the constraints on epipolar geometries in a view triplet and construct a new minimal representation of three views that guarantees consistency among all pairwise relations.

Chapter 3

Minimal Trifocal Model

This chapter presents a new minimal projective model of view triplets that enforces trifocal constraints on all pairwise epipolar geometries. The trifocally constrained collection of epipolar geometries performs the same role as the trifocal tensor when all three cameras lie in general position. Both models encode information about the 3D projective relations among views. However, the three pairwise epipolar geometry estimates express constraints on point pairs as well as on point triplets, whereas the trifocal tensor requires point or line triplets to be matched across three views. Since finding reliable matches across two views is easier, there is often more data available for accurately estimating pairwise geometries.

The trifocally constrained triplet model presented here is minimally parameterized, meaning there are no redundant degrees of freedom, and therefore the parameterization guarantees a globally consistent projective model of the three cameras. Like the trifocal tensor, it does not treat all three views symmetrically. Instead, it constrains one view pair's relation using the other two view pairs' relations. When two views in a collection have insufficient view overlap, their matching image points do not provide stable estimates of the relative camera geometry. By constraining their relation to satisfy trifocal constraints with respect to a third view, a more accurate, geometrically consistent estimate is possible. Such a scenario is common for image collections covering an extended scene.

The minimal model represents the trifocal geometry in terms of what we will call the *epipolar parameters*—the minimally parameterized epipoles and the epipolar collineations—between all three camera pairs. By construction, every point in the minimal parameter

space corresponds to a geometrically feasible triplet of projective camera configurations. We explicitly model the geometric dependencies among the epipoles and epipolar collineations between all pairs in a view triplet to obtain a minimal 18-parameter representation of the three-view configuration. The construction of the minimal parameter space dictates an incremental optimization technique for estimating trifocal geometry: first, two independent minimal 7-parameter epipolar geometries are fit to matching point pairs measured across two of the view pairs. Next, the 4 parameters of the third pair's dependent epipolar geometry are fit to point matches across the third pair. When there is asymmetry in the available image point matches between view pairs, this minimal model offers improved accuracy in the epipolar geometry estimates, and consequently improves the initial estimates of Euclidean rotation and translation that are needed for any application requiring initial knowledge of the 3D camera poses.

3.1 Trifocally Dependent Epipolar Parameters

When three cameras have optical centers in general position in space, the trifocal geometry is completely described by the three sets of consistent pairwise epipolar geometries. As shown in Section 2.3.3 (Table 2.1), the dependencies between the three epipolar geometries may be described algebraically in terms of the fundamental matrices. We now construct the geometric dependencies among the three epipolar geometries and make several key observations that will motivate the trifocally constrained model and the procedure for minimally parameterizing a view triplet.

Suppose that the epipolar geometries of view pairs (i, j) and (i, k) are known. These two relations are independent in the absence of any other information. What can be said about the epipolar geometry of view pair (j, k) given these two known relations?

Recall that the trifocal plane is the unique plane containing all three optical centers, and its projection into each image is a trifocal line (Figure 3-1). The trifocal lines are the structures that bind all three epipolar geometries. The basic observation that links the epipolar and trifocal geometry is that the trifocal lines are epipolar lines for all three view pairs. To see this for view pair (j, k), simply treat the optical center $\mathbf{C}_i$ of view i as an

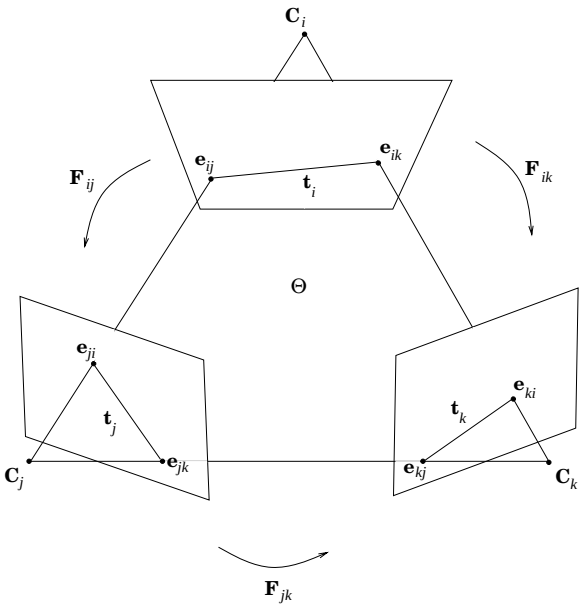


Figure 3-1: *The trifocal plane* $\mathbf{\Theta}$ *is defined by the three optical centers,* $< \mathbf{C}_i, \mathbf{C}_j, \mathbf{C}_k >$. *Each trifocal line is the projection of the trifocal plane into the image plane. It is also an epipolar line for each of the two other views.*

arbitrary scene point in space. Then the trifocal plane is an epipolar plane for the pair (j, k), and its projection into each image, $\mathbf{t}_j$ and $\mathbf{t}_k$, is an epipolar line. This brings us to the first observation.

Observation 1 *Given a view triplet, the two epipoles in any one view must lie on the trifocal line in that view.*

This is clear from Figure 3-1. We express this observation algebraically for all three views by the forms,

$$\mathbf{t}_i \simeq \mathbf{e}_{ij} \times \mathbf{e}_{ik} \tag{3.1}$$

$$\mathbf{t}_j \simeq \mathbf{e}_{ji} \times \mathbf{e}_{jk} \tag{3.2}$$

$$\mathbf{t}_k \simeq \mathbf{e}_{ki} \times \mathbf{e}_{kj}. \tag{3.3}$$

The second observation again treats the trifocal lines as epipolar lines.

Observation 2 *Given a view triplet, any two trifocal lines are images of each other under the epipolar collineation, the transformation of the epipolar pencil.*

For now, we will simply state this observation abstractly[1] for each view pair as

$$\mathbf{t}_i \simeq h_{ji}(\mathbf{t}_j) \tag{3.4}$$
$$\mathbf{t}_i \simeq h_{ki}(\mathbf{t}_k) \tag{3.5}$$
$$\mathbf{t}_j \simeq h_{ij}(\mathbf{t}_i) \tag{3.6}$$
$$\mathbf{t}_j \simeq h_{kj}(\mathbf{t}_k) \tag{3.7}$$
$$\mathbf{t}_k \simeq h_{ik}(\mathbf{t}_i) \tag{3.8}$$
$$\mathbf{t}_k \simeq h_{jk}(\mathbf{t}_j) \tag{3.9}$$
$$\tag{3.10}$$

.

We make one final observation that will be useful in practice when fitting parameters to view triplets.

Observation 3 *Given only the epipolar geometries between 2 pairs, (i, j) and (i, k) of a view triplet, the trifocal lines in views j and k are completely determined.*

We state this using the simple property that the fundamental matrix maps a point in one image to its matching point's epipolar line in the other image. In Figure 3-1, $\mathbf{e}_{ik}$ and $\mathbf{e}_{jk}$ may be treated as a matching point pair since they are both images of the point $\mathbf{C}_k$, so the fundamental matrix $\mathbf{F}_{ij}$ maps $\mathbf{e}_{ik}$ to its epipolar line in view j, which is simply the

[1] Recall the notation for the epipolar collineation:

$h_{jk}(\cdot)$	abstract function that maps an epipolar line from view j to view k
$\mathbf{H}_{jk}$	3×3 matrix representing the projective linear transformation that maps the 2D projective lines $\mathbf{l}_j \in \mathbb{P}^2$ to $\mathbf{l}_k \in \mathbb{P}^2$
$\mathbf{H}_{jk_{2\times 2}} \simeq \begin{bmatrix} a & b \\ c & d \end{bmatrix}$	2×2 matrix representing the projective linear transformation that maps a line in the epipolar pencil in view j, represented as a point $\mathbf{x}_{\mathbf{l}_j}$ on a projective line to a line in the epipolar pencil in view k, represented as a point $\mathbf{x}_{\mathbf{l}_k}$ on a projective line
$\mathbf{h}_{jk} \simeq [a, b, c, d]^T$	unrolled vector form of $\mathbf{H}_{jk_{2\times 2}}$

trifocal line $\mathbf{t}_j$. Stating this for all permutations of i, j, and k, we obtain,

$$\mathbf{t}_i \simeq \mathbf{F}_{ji}\mathbf{e}_{jk} \tag{3.11}$$

$$\mathbf{t}_i \simeq \mathbf{F}_{ki}\mathbf{e}_{kj} \tag{3.12}$$

$$\mathbf{t}_j \simeq \mathbf{F}_{ij}\mathbf{e}_{ik} \tag{3.13}$$

$$\mathbf{t}_j \simeq \mathbf{F}_{kj}\mathbf{e}_{ki} \tag{3.14}$$

$$\mathbf{t}_k \simeq \mathbf{F}_{ik}\mathbf{e}_{ij} \tag{3.15}$$

$$\mathbf{t}_k \simeq \mathbf{F}_{jk}\mathbf{e}_{ji}. \tag{3.16}$$

We have thus derived the geometric dependencies among the three epipolar geometries in a view triplet. The next section uses these relations as constraints on the epipolar parameters, and by explicitly modeling those constraints, constructs a minimal parameterization of three views.

3.2 Construction of the Minimal Parameters

This section constructs a minimal set of parameters that describes the trifocal geometry using the dependent epipolar parameters between all the pairs in a view triplet. Suppose we are given three views, $\{i, j, k\}$, and let $\mathbf{p}^7_{ij}$, $\mathbf{p}^7_{ik}$, and $\mathbf{p}^7_{jk}$ be the epipolar parameters between each view pair, as defined in Equation 2.40. The 21-parameter union of these three sets must have dependent elements since a view triplet has only 18 projective degrees of freedom. By explicitly enforcing the geometric constraints among these parameters in terms of the trifocal lines, we will construct a minimal 18-parameter projective model of three views. The approach is to partition the full set of epipolar parameters into two sets, one set of independent variables and one set of dependent variables. The dependent variables will be directly computable from the independent variables, and the union of the two sets will form a geometrically consistent set of epipolar parameters.

Suppose that the epipolar geometry for two view pairs (i, j) and (i, k) is known, so we have established the parameters $\mathbf{p}^7_{ij}$ and $\mathbf{p}^7_{ik}$. In the absence of all other information, these two parameter sets are independent, so they account for 14 degrees of freedom in the camera

configuration. However, since three views have a total of 18 projective degrees of freedom (Section 2.5), there are 4 degrees of freedom remaining to describe the epipolar relation between views j and k.

We seek a parameterization $\mathbf{p}_{jk}^4$ to express this relation. The key insight to defining the full epipolar geometry with only 4 parameters is to first construct trifocal lines in views j and k using Observation 3:

$$\begin{aligned} \mathbf{t}_j &\simeq \mathbf{F}_{ij}\mathbf{e}_{ik} \\ \mathbf{t}_k &\simeq \mathbf{F}_{ik}\mathbf{e}_{ij} \end{aligned}$$

The fundamental matrices $\mathbf{F}_{ij}$ and $\mathbf{F}_{ik}$ are directly computable from the minimal parameter sets $\mathbf{p}_{ij}^7$ and $\mathbf{p}_{ik}^7$ using the construction in Equation 2.28.

Now that the trifocal lines are known, Observation 1 requires that $\mathbf{e}_{jk}$ and $\mathbf{e}_{kj}$ must lie on $\mathbf{t}_j$ and $\mathbf{t}_k$ respectively, so each epipole is only free to vary along its trifocal line. We use this fact to express each unknown epipole by a single projective parameter on the trifocal line.

Choose two points on $\mathbf{t}_j$ and express the epipole $\mathbf{e}_{jk}$ as their linear combination. If the coordinates of the trifocal line in view j are $\mathbf{t}_j = [t_{j1}, t_{j2}, t_{j3}]^T$, then it must contain the points,

$$\mathbf{x}_j = [0, -t_{j3}, t_{j2}]^T \tag{3.17}$$

$$\mathbf{y}_j = [-t_{j3}, 0, t_{j1}]^T. \tag{3.18}$$

Therefore, we can express the epipole by the single parameter γ_j such that

$$\mathbf{e}_{jk} \simeq \gamma_j \mathbf{x}_j + \mathbf{y}_j. \tag{3.19}$$

Likewise, the trifocal line $\mathbf{t}_k = [t_{k1}, t_{k2}, t_{k3}]^T$ contains the points,

$$\mathbf{x}_k = [0, -t_{k3}, t_{k2}]^T \tag{3.20}$$

$$\mathbf{y}_k = [-t_{k3}, 0, t_{k1}]^T. \tag{3.21}$$

so the epipole $\mathbf{e}_{kj}$ may be parameterized by the single parameter γ_k such that

$$\mathbf{e}_{kj} \simeq \gamma_k \mathbf{x}_k + \mathbf{y}_k. \tag{3.22}$$

For any γ_j (resp. γ_k), we can reconstruct the coordinates of the epipole, given the trifocal line $\mathbf{t}_j$ (resp. $\mathbf{t}_k$). This construction, depicted in Figure 3-2, will be used in the procedure given in Chapter 4 for estimating the single epipole parameter along a known trifocal line.

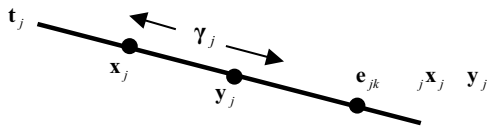


Figure 3-2: *The epipole* $\mathbf{e}_{jk}$ *lies on the trifocal line* $\mathbf{t}_j$*, parameterized by the scalar* γ_j*.*

There are two remaining degrees of freedom left to describe the epipolar geometry for view pair (j, k). The epipolar collineation h_{jk}, which in general has 3 degrees of freedom, must therefore be constrained to have only 2 free parameters. According to Observation 2, the trifocal line must satisfy

$$\mathbf{t}_k \simeq h_{jk}(\mathbf{t}_j).$$

This constraint treats the trifocal line as an epipolar line and therefore as a member of the epipolar pencil. It gives us a single match for the epipolar collineation, depicted in Figure 3-3.

To use the trifocal line match as an explicit constraint on the parameters that define the epipolar collineation, we will use the same construction as in Section 2.2.4. Represent the epipolar collineation as a transformation of the projective line by defining a transverse line in each view. As depicted in Figure 3-4, we let $< \mathbf{p}_j, \mathbf{q}_j >$ and $< \mathbf{p}_k, \mathbf{q}_k >$ be the transverse lines in views j and k, respectively, and write each transverse line's intersection with the trifocal line as a point on a projective line,

$$\mathbf{x}_{\mathbf{t}_j} \simeq [\alpha_j, \beta_j]^T$$

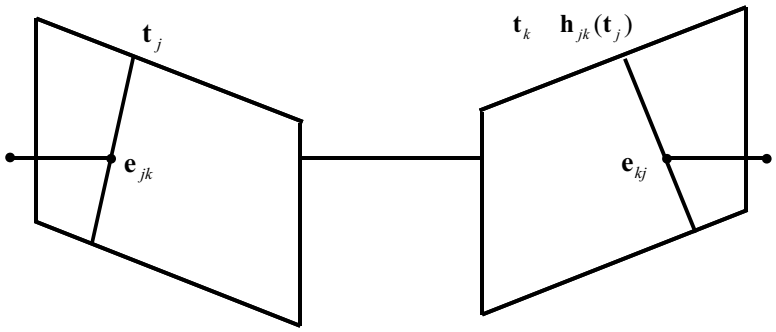


Figure 3-3: *The trifocal lines* $\mathbf{t}_j$ *and* $\mathbf{t}_k$ *are in correspondence via the epipolar collineation.*

$$\mathbf{x}_{\mathbf{t}_k} \simeq [\alpha_k, \beta_k]^T,$$

where Equations 2.24 to 2.27 now become

$$\alpha_j = \mathbf{t}_j^T \mathbf{q}_j \tag{3.23}$$

$$\beta_j = -\mathbf{t}_j^T \mathbf{p}_j \tag{3.24}$$

$$\alpha_k = \mathbf{t}_k^T \mathbf{q}_k \tag{3.25}$$

$$\beta_k = -\mathbf{t}_k^T \mathbf{p}_k. \tag{3.26}$$

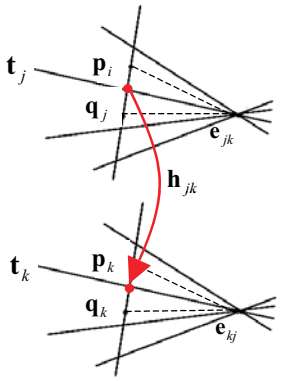


Figure 3-4: *The trifocal lines* $\mathbf{t}_j$ *and* $\mathbf{t}_k$*, each represented as an epipolar line parameterized by two basis vectors, are in correspondence via the epipolar collineation.*

The projective representations of the trifocal lines are related by the collineation parameters

$$\mathbf{h}_{jk} \simeq [a, b, c, d]^T,$$

via Equation 2.23,

$$a(\alpha_j\beta_k) + b(\beta_j\beta_k) - c(\alpha_j\alpha_k) - d(\beta_j\alpha_k) = 0.$$

Defining the vector of coefficients

$$\mathbf{k}^T = [\alpha_j\beta_k, \beta_j\beta_k, -\alpha_j\alpha_k, -\beta_j\alpha_k]^T, \tag{3.27}$$

Equation 2.23 now becomes

$$\mathbf{k}^T\mathbf{h}_{jk} = 0, \tag{3.28}$$

a single linear equation in the components of the collineation.

To constrain the collineation to only 2 parameters, first choose a normalization index, c_4, and normalize the collineation components so that $\mathbf{h}_{jk}(c_4) = 1$. Now choose an index c_3 to be the dependent parameter, $\mathbf{h}_{jk}(c_3)$. Express the dependent parameter in terms of the two remaining ones, by computing the α's and β's for the known trifocal lines using Equation 3.23, and substituting these into Equation 3.28. Given the two collineation components, $\mathbf{h}_{jk}(c_1)$ and $\mathbf{h}_{jk}(c_2)$, the remaining component is completely determined by isolating $\mathbf{h}_{jk}(c_3)$ in Equation 3.28:

$$\mathbf{h}_{jk}(c_3) = -\frac{\mathbf{h}_{jk}(c_1)\mathbf{k}(c_1) + \mathbf{h}_{jk}(c_2)\mathbf{k}(c_2) + \mathbf{k}(c_4)}{\mathbf{k}(c_3)}. \tag{3.29}$$

To ensure that this parameter is well-defined, it is necessary to examine the parameterized trifocal line and choose the index set $\zeta_c = [c_1, c_2, c_3, c_4]^T$ such that the c_3^{th} component of the coefficient vector $\mathbf{k}$ is nonzero.

We have now parameterized the epipolar collineation by the two values $\mathbf{h}_{jk}(c_1)$ and $\mathbf{h}_{jk}(c_2)$. These parameters together with the epipole parameters γ_j and γ_k define the set of

4 parameters defining the geometry of the view pair (j, k):

$$\mathbf{p}_{jk}^{4} \quad = \quad \{\gamma_j, \gamma_k, \mathbf{h}_{jk}(c_1), \mathbf{h}_{jk}(c_2)\}. \tag{3.30}$$

From this set, $\mathbf{p}_{jk}^{7}$ is computable by using Equations 3.19 and 3.22 to construct the epipoles $\mathbf{e}_{jk}$ and $\mathbf{e}_{kj}$, and then using Equation 3.29 to define $\mathbf{h}_{jk}(c3)$ and setting $\mathbf{h}_{jk}(c4) = 1$. We now have a technique for minimally representing the projective relations among three views as the union of two independent sets of epipolar parameters with one dependent epipolar parameter set:

$$\mathbf{p}_{ijk}^{18} \quad = \quad \mathbf{p}_{ij}^{7} \cup \mathbf{p}_{ik}^{7} \cup \mathbf{p}_{jk}^{4}. \tag{3.31}$$

3.3 Comparison to the Trifocal Tensor

This section briefly discusses the relationship between the trifocally constrained epipolar geometries and the well-known trifocal tensor for projective modeling of view triplets.

Assume for now that the view triplet is not collinear, and suppose a scene point that does not lie in the trifocal plane is matched across all three views. The trifocally constrained epipolar geometries are sufficient to enforce the constraint that all three optical rays meet at a single scene point. Using only the epipoles and the epipolar collineations, we can write down the requirement that each image point in a matching triplet must lie at the intersection of the two epipolar lines mapped from the other two image points into the current view. This statement is a view-based expression of the triangulation constraint, that the three optical rays through three imaged points must intersect in a single 3D point, the same constraint expressed by the trifocal tensor. Consequently, the point transfer property of trifocal tensors is also expressible in terms of the epipoles and collineations.

If the three sets of epipolar parameters are not consistent, then they do not properly enforce the trifocal constraints on image points because they do not describe a physically plausible camera configuration. Similarly, if a $3 \times 3 \times 3$ trifocal tensor is not properly parameterized or if its coefficients are not forced to meet certain nonlinear constraints, then it does not necessarily correspond to a unique projective camera configuration. From an

applications perspective, the trifocal tensor is often used as an initial step for bundle adjustment or structure from motion, but the typical method for recovering a set of projection matrices from the trifocal tensor requires extracting the epipoles from the trifocal tensor to construct projection matrices. The minimal epipolar parameter set bypasses the tensor computation and directly computes the epipolar geometries. It would be interesting to compare the epipole accuracy obtained from the trifocal tensor to the estimate found by directly fitting the trifocally constrained epipolar parameters.

In practice, the trifocal tensor is only recoverable when there are sufficient point matches across all three views to constrain the trilinear equations. Often, there are many more pairs of point matches measured between images than triplets. The trifocally constrained epipolar parameter set allows the view triplet to be analyzed from only matching point pairs while guaranteeing a legitimate camera configuration.

A degenerate configuration for the triplet of dependent epipolar parameters is the case of collinear optical centers. Suggestions for detecting this case are discussed in Section 5.3 and Chapter 6.

3.4 Summary

This chapter has presented a model for the projective geometry of view triplets when the optical centers are in general position. The geometric dependencies between all three pairwise epipolar geometries in a view triplet were derived, and a new, minimally parameterized, consistent set of epipolar parameters among three views was defined. The next chapter presents an algorithm for estimating this model, and gives empirical examples of its improved accuracy for estimating the geometry of view triplets.

Chapter 4

Trifocally Constrained Parameter Estimation

This chapter describes a procedure for estimating a minimal trifocal model for a view triplet by imposing trifocal constraints on the triplet of epipolar geometries. The algorithm initializes the pairwise epipolar geometries of all three view pairs in a triplet using existing linear and nonlinear methods. It then projects the initialized model onto the space of trifocally constrained epipolar parameters, and performs an optimization in the minimal space. Experimental results on image sequences whose image pairs have varying overlaps in fields of view show that the trifocal constraints correct errors in estimates of epipolar geometry from relatively few point matches. Chapter 5 analyzes the behavior of the objective functions evaluated over the minimal epipolar parameter space and offers a generalized parameterization that handles sensitive camera configurations.

4.1 The 4-Parameter Algorithm

We now introduce a new minimally parameterized trifocal algorithm, which takes as input a set of matching image points between pairs of views in a triplet of images and returns a consistent set of epipolar parameters that describe the epipolar geometry between all three view pairs. By consistent, we mean that the pairwise epipolar geometries satisfy all the projective constraints described in Chapter 3, and hence correspond to a legal projective

camera configuration. The main procedure is the 4-parameter algorithm: given reliable estimates of the epipoles and epipolar collineations between two view pairs in a triplet, $\mathbf{p}_{ij}^7$ and $\mathbf{p}_{ik}^7$, it performs a nonlinear search for the optimal trifocally constrained epipolar parameter set, $\mathbf{p}_{jk}^4$. These three sets comprise the minimal projective parameterization of three views.

4.1.1 Initialization

To initialize the nonlinear search, obtain a starting point in the parameter space by using the 8-point algorithm to fit a linear least-squared estimate of the fundamental matrix to the matching image points in each view pair [19]. The 8-point algorithm returns the linear least-squares solution to the error, $\Sigma \mathbf{m}_j^T \mathbf{F}_{ij} \mathbf{m}_i$, summed over all matching image points. The RANSAC algorithm [15] for robust sampling is used to eliminate outliers from the set of correspondences.

4.1.2 Extraction

For all view pairs, extract the epipolar parameters from the fundamental matrix using the canonical decomposition described in Section 2.2.4. Choosing a set of of normalization indices, decompose the fundamental matrix $\mathbf{F}_{ij}$ into the epipolar parameters $\mathbf{p}_{ij}^7 = \{\mathbf{e}_{ij}(i_1), \mathbf{e}_{ij}(i_2), \mathbf{e}_{ji}(j_1), \mathbf{e}_{ji}(j_2), \mathbf{h}_{ij}(c_1), \mathbf{h}_{ij}(c_2), \mathbf{h}_{ij}(c_3)\}$, and likewise for $\mathbf{F}_{ik}$ and $\mathbf{F}_{jk}$.

For the two independent view pairs (i, j) and (i, k), perform a nonlinear optimization over each parameter set, $\mathbf{p}_{ij}^7$ and $\mathbf{p}_{ik}^7$, as described in [5]. Any objective function may be used that expresses a constraint on matching image points in terms of projective relations. We use the distance from each measured point to its epipolar line under the current model of the epipolar geometry [53]. Details of this objective function are given in Section 4.1.4 and discussed further in Chapter 5. The function is only evaluated on the robust sample of points found by the 8-point algorithm. A standard nonlinear least-squares optimization algorithm such as Levenberg-Marquardt is used.

4.1.3 Projection

Fix the optimized parameters, $\mathbf{p}_{ij}^7$ and $\mathbf{p}_{ik}^7$, and construct the minimal set of trifocally constrained epipolar parameters $\mathbf{p}_{jk}^4 = \{\gamma_j, \gamma_k, \mathbf{h}_{jk}(c_1), \mathbf{h}_{jk}(c_2)\}$ in Equation 3.30 from $\mathbf{F}_{jk}$ as follows. First, define the initial epipoles to be the kernel of $\mathbf{F}_{jk}$ and project them onto the trifocal lines to find the parameters γ_j and γ_k that satisfy Equations 3.19 and 3.22. Then construct the collineation parameters by first extracting the collineation $\mathbf{h}_{jk}$ directly from the components of $\mathbf{F}_{jk}$ as shown in the example in Section 2.2.4. Using the chosen normalization indices, keep only the two collineation values $\mathbf{h}_{jk}(c_1)$ and $\mathbf{h}_{jk}(c_2)$ to include in the 4-parameter projection.

4.1.4 Optimization

Using the projected parameters $\mathbf{p}_{jk}^4$ as an initial point, perform a nonlinear least-squares optimization of the 4D parameter space by minimizing an objective function that measures the fit of the current parameters to the measured image point correspondences.

The typical 2-view error function for estimating epipolar geometry sums the weighted Euclidean distance from each image point in the data set to the epipolar line generated from its matching point by the current epipolar geometry model. Let $(\mathbf{m}_j, \mathbf{m}_k)$ be a matching pair of points. The squared *point-to-epipolar-line* error for a point pair given the parameters is defined as

$$e_l^2(\mathbf{m}_j, \mathbf{m}_k; \mathbf{p}_{jk}^4, \mathbf{p}_{ij}^7, \mathbf{p}_{ik}^7) \quad = \quad d_l^2(\mathbf{m}_j, \mathbf{l}_j) + d_l^2(\mathbf{m}_k, \mathbf{l}_k), \tag{4.1}$$

where the epipolar lines are

$$\begin{aligned} \mathbf{l}_j &\simeq h_{kj}(\mathbf{e}_{kj} \times \mathbf{m}_k) \simeq \mathbf{F}_{jk}^T \mathbf{m}_k \simeq (l_1, l_2, l_3)^T \\ \mathbf{l}_k &\simeq h_{jk}(\mathbf{e}_{jk} \times \mathbf{m}_j) \simeq \mathbf{F}_{jk} \mathbf{m}_j \simeq (l'_1, l'_2, l'_3)^T \end{aligned}$$

and the point-line distance is defined as [53]

$$d_l^2(\mathbf{m}_j, \mathbf{l}_j) \quad = \quad \frac{1}{\sqrt{l_1^2 + l_2^2}} (\mathbf{m}_j^T \mathbf{l}_j)^2$$

$$d_l^2(\mathbf{m}_k, \mathbf{l}_k) = \frac{1}{\sqrt{{l'_1}^2 + {l'_2}^2}}(\mathbf{m}_k^T \mathbf{l}_k)^2.$$

For this to define the Euclidean distance, the points $\mathbf{m}_j$ and $\mathbf{m}_k$ are embedded in $\mathbb{R}^3$ by normalizing them by their third components.

At each iteration, the epipoles $\mathbf{e}_{jk}$ and $\mathbf{e}_{kj}$ and the collineation h_{jk} are constructed from $\mathbf{p}_{jk}^4$ using the construction step given in the next section. The epipolar line $\mathbf{l}_j$ may be represented as a point on the projective line, using Equations 2.18 and 2.19, and then mapped to $\mathbf{l}_k$ using $\mathbf{H}_{jk_{2\times 2}}$ and Equation 2.21. Alternatively, the fundamental matrix $\mathbf{F}_{jk}$ may be constructed from the epipolar parameter set and applied directly to the projective line $\mathbf{l}_j$. These are theoretically equivalent since the fundamental matrix is directly constructed from the epipolar parameters. The objective function is the sum of this error over all point matches between views j and k.

An alternate objective function may be used that operates on image point triplets and represents the true reprojection error, that is, the distance from each measured image point to its predicted position given the current model. The predicted image point in a view is the intersection of the two estimated epipolar lines projected from the matching points in the other two images:

$$e_p^2(\mathbf{m}_j, \mathbf{m}_k; \mathbf{p}_{jk}^4, \mathbf{p}_{ij}^7, \mathbf{p}_{ik}^7, \mathbf{m}_i) = d_p^2(\mathbf{m}_j, \mathbf{l}_{j_{\mathbf{m}_i}} \times \mathbf{l}_{j_{\mathbf{m}_k}}) + d_p^2(\mathbf{m}_k, \mathbf{l}_{k_{\mathbf{m}_i}} \times \mathbf{l}_{k_{\mathbf{m}_j}}), \quad (4.2)$$

where d_p is the Euclidean distance between points, and the epipolar lines may be computed from either the collineations or the fundamental matrices:

$$\begin{aligned}
\mathbf{l}_{j_{\mathbf{m}_i}} &\simeq h_{ij}(\mathbf{e}_{ij} \times \mathbf{m}_i) \simeq \mathbf{F}_{ij}\mathbf{m}_i \\
\mathbf{l}_{j_{\mathbf{m}_k}} &\simeq h_{kj}(\mathbf{e}_{kj} \times \mathbf{m}_k) \simeq \mathbf{F}_{kj}\mathbf{m}_k \\
\mathbf{l}_{k_{\mathbf{m}_i}} &\simeq h_{ik}(\mathbf{e}_{ik} \times \mathbf{m}_i) \simeq \mathbf{F}_{ik}\mathbf{m}_i \\
\mathbf{l}_{k_{\mathbf{m}_j}} &\simeq h_{jk}(\mathbf{e}_{jk} \times \mathbf{m}_j) \simeq \mathbf{F}_{jk}\mathbf{m}_j.
\end{aligned}$$

The disadvantage of this objective function is that it requires each point to be matched across all three views, just as the trifocal tensor does, rather than using pairwise point

matches which are in practice more readily available. Section 5.1 discusses several alternative objective functions.

4.1.5 Reconstruction

To reconstruct $\mathbf{p}^7_{jk}$ from these $\mathbf{p}^4_{jk}$, first compute the epipoles from γ_j and γ_k using Equations 3.19 and 3.22. Then reconstruct the collineation by setting $\mathbf{h}_{jk}(c_4) = 1$, and defining the coefficients $\mathbf{k}$ using Equation 3.27 and Equations 3.23. Finally, compute the collineation parameter $\mathbf{h}_{jk}(c_3)$ from Equation 3.29. We have now reconstructed $\mathbf{p}^7_{jk}$, a set of epipolar parameters relating views j and k that is consistent with the other two sets of epipolar parameters, $\mathbf{p}^7_{ij}$ and $\mathbf{p}^7_{ik}$. The epipoles, epipolar collineations, fundamental matrices, and trifocal lines constructed from this parameter set will satisfy all the constraints described in Chapter 3.

4.2 Projective to Euclidean Conversion

The experiments in this chapter will compare the accuracy of projective parameter estimates in the projective space as well as in terms of their physical effects on cameras. This is done by mapping the epipolar parameters onto a set of Euclidean parameters using the intrinsic parameters from in each simulation. Due to constraints on Euclidean quantities that do not exist for projective quantities, such as the orthonormality of rotation matrices, this projection is many-to-one. Here, we briefly discuss the issues surrounding this mapping and refer to the more detailed analyses in the literature [32, 52].

In projective spaces, only incidence properties are defined, which may also thought of as unions and intersections of projective subspaces, *i.e.*, two points in a plane define a line, three planes in space intersect at a point, *etc.*. There is no metric defined on the projective space and hence no sense of orthogonality. To move to a Euclidean coordinate frame, the internal camera parameters are needed.

4.2.1 The Essential Matrix

The perspective projection equation in Equation 2.1 and the factored camera projection matrix given in Equation 2.3 give us a factored representation of the pinhole camera model:

$$\mathbf{m} \simeq \mathbf{A}[\mathbf{R}\ \mathbf{T}]\mathbf{M}. \tag{4.3}$$

Suppose the internal camera parameters are known for two views i and j related by the unknown rotation $\mathbf{R}$ and unknown translation $\mathbf{T}$, so a scene point $\mathbf{M}$, written in the coordinate frame of view j, is located at the point $\mathbf{RM}+\mathbf{T}$ in the coordinate frame of view i. The matrices $\mathbf{A}_i$ and $\mathbf{A}_j$ may be constructed from the internal parameters, and the two images $\mathbf{m}_i$ and $\mathbf{m}_j$ of $\mathbf{M}$ may be calibrated by inverting the effects of the internal parameters. Let $\mathbf{m'}_i$ and $\mathbf{m'}_j$ be the two calibrated image points,

$$\begin{aligned} \mathbf{m'}_i &\simeq \mathbf{A}_i^{-1}\mathbf{m}_i \\ \mathbf{m'}_j &\simeq \mathbf{A}_j^{-1}\mathbf{m}_j. \end{aligned}$$

Working in calibrated coordinates, it is easy to show that the coplanarity constraint between two views reduces takes the same form as in the uncalibrated case, except that now two matching points are related by the matrix $\mathbf{E}_{ij}$ such that

$$\mathbf{m'}_j\mathbf{E}_{ij}\mathbf{m'}_i = \mathbf{0},$$

and

$$\mathbf{E}_{ij} \simeq [\mathbf{T}]_\times\mathbf{R}. \tag{4.4}$$

This is called the *essential matrix*, due to Longuet-Higgins [25], and its relation to the fundamental matrix is [11, 17],

$$\mathbf{E}_{ij} \simeq \mathbf{A}_j^T\mathbf{F}_{ij}\mathbf{A}_i. \tag{4.5}$$

This equality gives us a way of mapping a fundamental matrix to the essential matrix given the internal parameters of both cameras. The essential matrix depends purely on the relative 3D rotation and translation between the cameras, so obtaining a Euclidean interpretation of the camera configuration is now a matter of factoring $\mathbf{E}_{ij}$ into $\mathbf{R}$ and $\mathbf{T}$. We treat the essential matrix as a projective quantity defined only up to scale, so any scalar multiple of $\mathbf{T}$ will satisfy the homogeneous equality in Equation 4.4. Therefore, it is only possible to recover the direction of the translation from the essential matrix.

There are in fact two possible rotations that satisfy Equation 4.4. These two rotations, combined with the two possible signs for the direction of translation give a total of 4 rotation-translation pairs that correspond to a given essential matrix. To disambiguate between the 4 choices of rotation and translation direction, each configuration may be tested using one image correspondence by projecting an optical ray outward from each camera center and reconstructing the 3D scene point location. The reconstructed scene point will only be in front of both cameras for one of the 4 choices.

Horn gives an exact method for extracting the rotation pair and translation direction from a valid essential matrix [31]. In practice, however, an essential matrix that has been computed linearly from observed image points will not necessarily satisfy Equation 4.4. It is known that for $\mathbf{E}_{ij}$ to be factorizable into a skew-symmetric matrix and an orthonormal rotation matrix, it must have two nonzero singular values of equal magnitude and one zero singular value [25].

A popular method for projecting an essential matrix into the space of admissible solutions is to adjust its singular values to satisfy these constraints. Setting the smallest singular value to 0 and forcing the 2 nonzero singular values to have equal magnitudes guarantees that the essential matrix is associated with a unique pair of legal rotations and two opposite translation directions. While this projection does not guarantee that the resulting essential matrix is the optimal solution for the original image data, it is often used with success in practice.

4.2.2 Self-Calibration

Equation 4.5 offers a way to compute the essential matrix from the internal parameters, but when working with uncalibrated imagery, these are not provided, and the original camera rig is not available for off-line calibration with a known 3D pattern. *Self-calibration* or *auto-calibration* is the problem of recovering the unknown internal camera patterns using only the projective relationships between views. It has been shown that under certain conditions, it is possible to recover the internal camera parameters by solving the Kruppa equations from the fundamental matrices [10, 26]. For a complete review of self-calibration methods, see [21].

4.2.3 Scaling Baselines

Given more than two views, the translation directions recovered from the pairwise essential matrices may be simultaneously scaled to obtain a Euclidean representation of the camera configuration defined up to a single unknown global scale factor. The multiple-view baseline scaling algorithm given in [1] finds the set of pairwise scale factors that best satisfies the constraint that all scaled baselines intersect at the optical centers.

4.2.4 Summary

The experiments in this work use the SVD projection of the essential matrix and the multiple-view baseline scaling algorithm to recover Euclidean interpretations of the estimated projective parameters. For the purposes of comparing methods, the synthetic internal parameters are used as nominal values in order to visualize the physical effects of errors in estimation. For practical applications, the internal parameters would need to be estimated using some form of self-calibration. The resulting camera configuration provides a starting point for Euclidean structure from motion, bundle adjustment, or any other application that requires initial knowledge of relative camera poses.

4.3 Experimental Results

This section presents results of the minimal trifocal parameterization on an image sequence that represents a common scenario. Due to asymmetry in the overlap between view pairs in the sequence, the number of available point matches varies greatly from one pair to another. We will demonstrate that the trifocal constraints can improve the accuracy of the estimated epipolar geometry between views having few measured point matches.

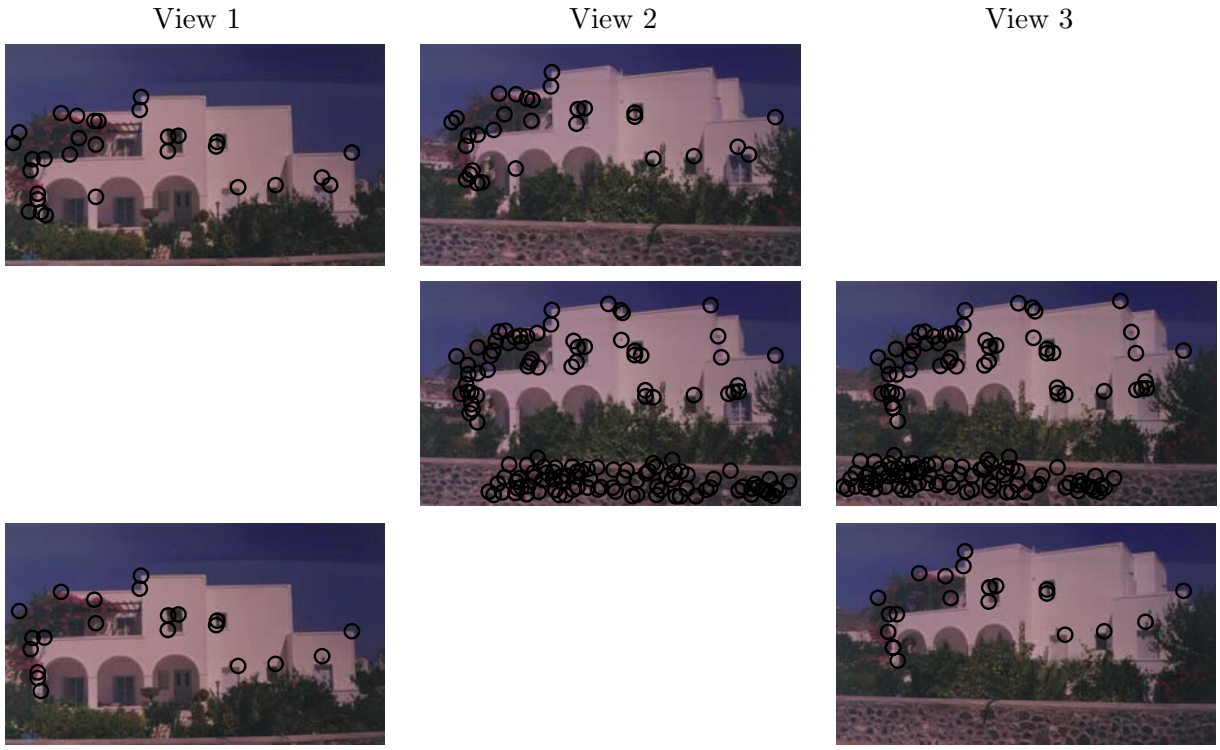


Figure 4-1: *Point correspondences for the* `santorini` *sequence. View pair (1,2) has 30 matches; pair (2,3) has 141 matches; pair (1,3) has 21 matches.*

To test how well our minimal model estimates the camera geometry of view pairs in a realistic imaging scenario, we compare it to two existing methods for determining epipolar geometry without enforcing trifocal constraints: the linear 8-point algorithm [19] and the nonlinear 7-parameter optimization [5], both described in 4.1.

The three sets of matching point pairs observed in the `santorini` image triplet shown in Figure 4-1 are used to construct the simulation. Image features are matched between all

3 view pairs using Zhang's image matching technique [55]. The image matcher may be run in several modes: the simplest mode uses standard correlation and relaxation techniques to match point pairs, and the most sophisticated mode interleaves the correlation with estimates of the epipolar geometry in order to discard false matches and constrain the correspondence search. We run the image matcher in its basic correlation mode in order to find as many points as possible, and then manually eliminate the many false point matches.

To perform controlled experiments, a synthetic ground truth for camera placement and scene structure is constructed by estimating the trifocally constrained epipolar geometries and projecting them to a Euclidean configuration. The synthetic Euclidean ground truth is computed by introducing a set of internal parameters with a focal length of 520 pixels, the principal point at the center of the image, and equally scaled, orthogonal image axes. The estimated epipolar geometry is projected to Euclidean rotations and translations using the standard method given in Section 4.2. The Euclidean scene point locations are constructed by triangulating the optical rays from these camera positions through the image point matches in each view [20], and the entire configuration is scaled, translated, and rotated to a realistic world coordinate frame. Using the original distributions of matching image points across frame pairs, we reproject the constructed scene points onto the images to obtain noise-free synthetic image points. Figure 4-2 shows an overhead view of the camera positions with the estimated scene points. Views 1 and 3 are the farthest apart and have the fewest matching points.

4.3.1 Projective Parameter Accuracy

The synthetic image points are corrupted by Gaussian noise with .5 pixel variance, and then three estimates are compared. All three epipolar geometries are initialized by the linear 8-point algorithm. The nonlinear 7-parameter optimization refines the estimates for view pairs (1,2) and (2,3), and the nonlinear 4-parameter optimization enforces the trifocal constraints on the final estimate for view pair (1,3). This estimate is compared to the estimates given by the 8-point algorithm alone, and by the 7-parameter optimization without enforcing trifocal constraints.

Figure 4-3 shows the resulting parameter values. The epipole plots are embedded in

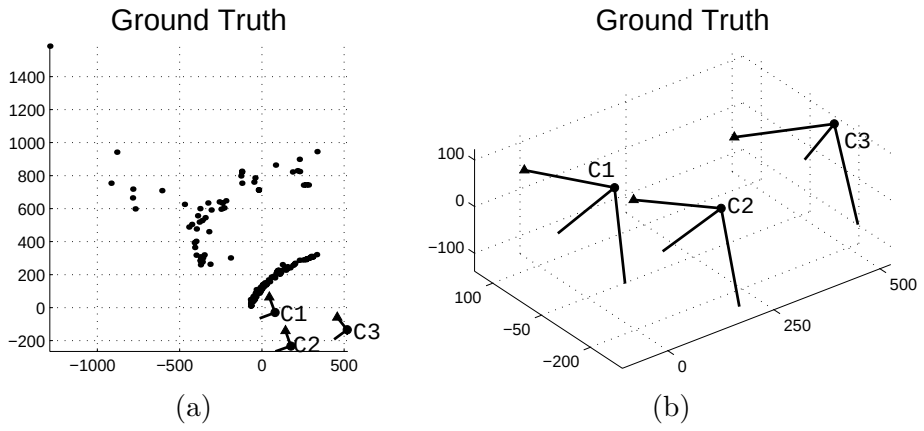


Figure 4-2: *Synthetic ground truth for* `santorini` *image sequence simulation. (a) Overhead view of cameras and reconstructed scene points. (b) Side view of camera poses.*
.

the Euclidean image plane. Note that they lie outside the boundaries of the 704×416 pixel image. The collineation plots are the three components of the 2×2 collineation matrix, scaled by the fourth component. Figure 4-3(c) shows the estimates projected into the trifocally constrained space.

Figure 4-4 shows the percent error in epipole estimation for the x and y components of the epipoles in both views. For all methods, the x-component error is much greater than the y-component error due to the well-known fact that epipole estimation has an anisotropic uncertainty for many common configurations. Chapter 5 explores in more detail the characteristics of the parameters and the objective function minimized over this parameter space. The error comparisons Figures 4-3 and 4-4 show that each projective parameter is significantly improved by the trifocal constraints, so fitting the epipolar geometry to a minimal representation can improve accuracy.

4.3.2 Euclidean Parameter Accuracy

To examine the effect that the accuracy of projective parameters has on the rigid Euclidean solution, we use the synthetic internal parameter matrix to project the epipoles and epipolar collineations to Euclidean interpretations. Figure 4-5 and Tables 4.1 and 4.2 show the errors in translations and rotations. Since the absolute scale of the configuration is not recoverable,

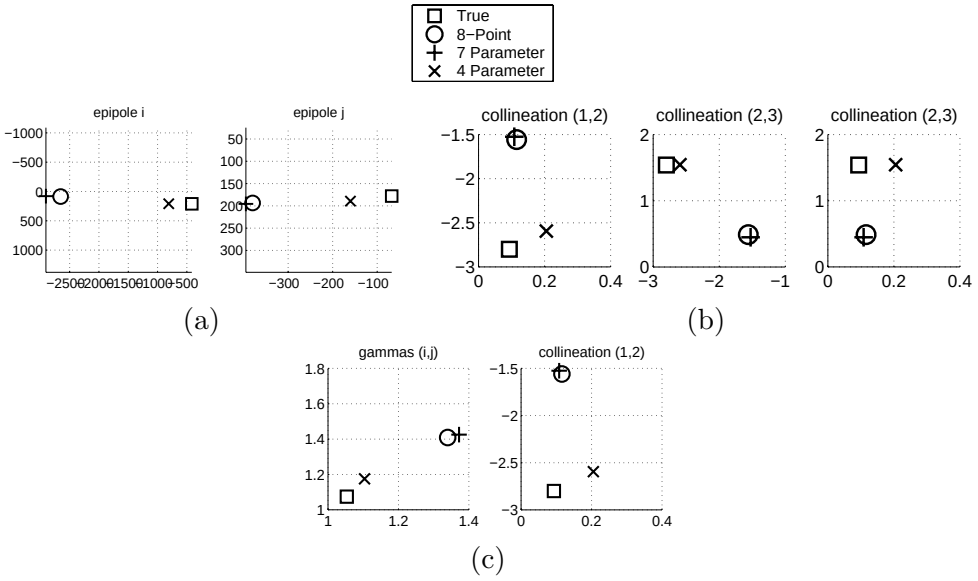


Figure 4-3: *Estimates of epipolar geometry parameters for views 1 and 3 of the* `santorini` *sequence simulation. (a) Components of reconstructed epipoles, embedded in Euclidean space and given in pixel coordinates. Left: epipole in view 1; right: epipole in view 3. (b) Three elements of reconstructed collineation matrix, defined up to scale. (c) Trifocal projection of each estimate. Left: the 2 epipole parameters; right: the 2 collineation parameters. The minimal trifocal parameterization significantly improves the accuracy of the projective parameters.*

the units of the translation error are arbitrary, and only relative values are meaningful. The minimal 4-parameter solution significantly improves the accuracy of the 3D camera poses.

Figure 4-6 shows the convergence paths for the `santorini` simulations. The epipole search is projected onto the trifocal line, which frees the search from the local minimum in the 7-dimensional search. The epipolar collineation is more difficult to visualize. Its 3 parameters are projected onto a 2D plane of collineations that preserve the trifocal line mapping.

The convergence paths show that by fixing the trifocal lines, the constrained parameter search is projected onto a lower-dimensional space of legal solutions that lead to the global minimum. The value of the objective function is initially high because the initial minimal projection has not yet been optimized and therefore does not necessary lie in a low-error region of the space. For the constrained parameterization to converge correctly, it is nec-

	Translation Error : Norm		Translation Error : Angle (Radians)		Rotation Error : Angle (Radians)	
	View 1	View 3	View 1	View 3	View 1	View 3
$8-$ Point	57.2	832	0.0045	0.2515	0.3014	0.3879
$7-$ Parameter	54.8	1006	0.0076	0.2594	0.0076	0.2594
$4-$ Parameter	21.1	130.0	0.0076	0.1247	0.1222	0.1358

Table 4.1: *Error in 3D camera translations and rotations for* `santorini` *sequence, projected from epipolar geometry estimates to Euclidean parameters using the simulated internal camera parameters. Each estimate's camera triplet is registered to the ground truth by alignment of the middle view.*

	Translation Error : X		Translation Error : Y		Translation Error : Z	
	View 1	View 3	View 1	View 3	View 1	View 3
$8-$ Point	26.4710	781.8983	50.6244	279.6619	3.3175	51.2033
$7-$ Parameter	26.5383	946.8517	47.8757	335.1389	3.4357	55.9343
$4-$ Parameter	12.5590	116.3677	16.9385	57.7711	1.2070	3.2653

Table 4.2: *Error in each component of the 3D camera translations for* `santorini` *sequence, projected from epipolar geometry estimates to Euclidean parameters using the simulated internal camera parameters. Each estimate's camera triplet is registered to the ground truth by alignment of the middle view.*

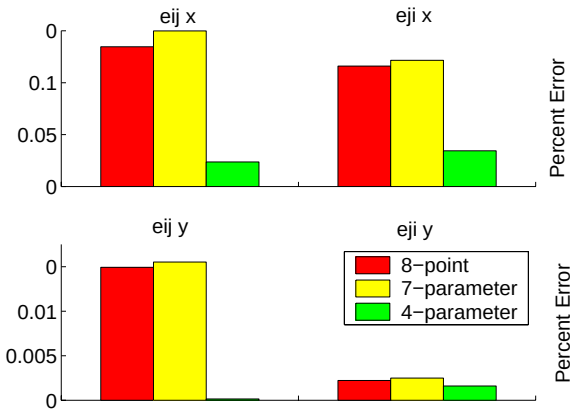


Figure 4-4: *Percent error of epipole estimation for views 1 and 3 of the* `santorini` *sequence. Top: x components of epipoles in each view. Bottom: y component of epipoles in each view. The 4-parameter trifocally constrained algorithm improves both components of both epipole estimates.*

essary that the 7D point representing the true solution lies near or within the 4D solution space. For a better understanding of these paths, Chapter 5 analyzes the local shape of the objective function at various minimima, and discusses the relationships between valleys in these surfaces and the geometry of camera configurations.

4.4 Summary

The experimental results in this chapter show how our minimally parameterized model of three views can provide a good starting point for structure from motion or bundle adjustment when three views of a scene have widely varying overlapping views. When the number of point matches between a view pair is small, the epipolar geometry estimation is generally less accurate. Our method overcomes this inaccuracy by enforcing consistency among all three pairwise geometries. The next chapter provides a detailed analysis of the numerical behavior of projective parameter optimization, and introduces a new parameterization with improved numerical stability.

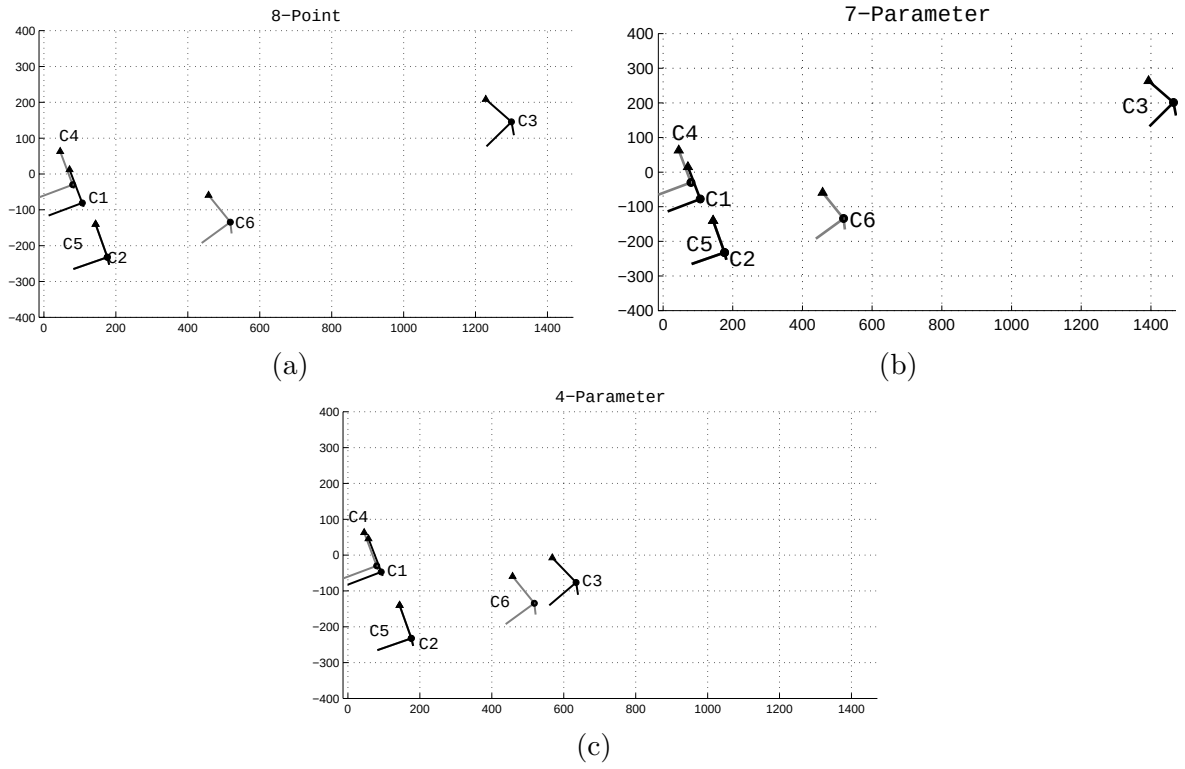


Figure 4-5: *Camera poses found by projecting epipolar geometry estimates to Euclidean rotations and translations for the* `santorini` *sequence simulation. Estimates are superimposed over ground truth. Camera labels (1,2,3) are estimates and camera labels (5,6,7) are the corresponding ground truth poses. Triplets are aligned with the middle view and units are in world coordinates. (a) Linear 8-point algorithm. (b) Nonlinear 7-parameter optimization without trifocal constraints. (c) Nonlinear 4-parameter optimization of trifocally constrained model.*

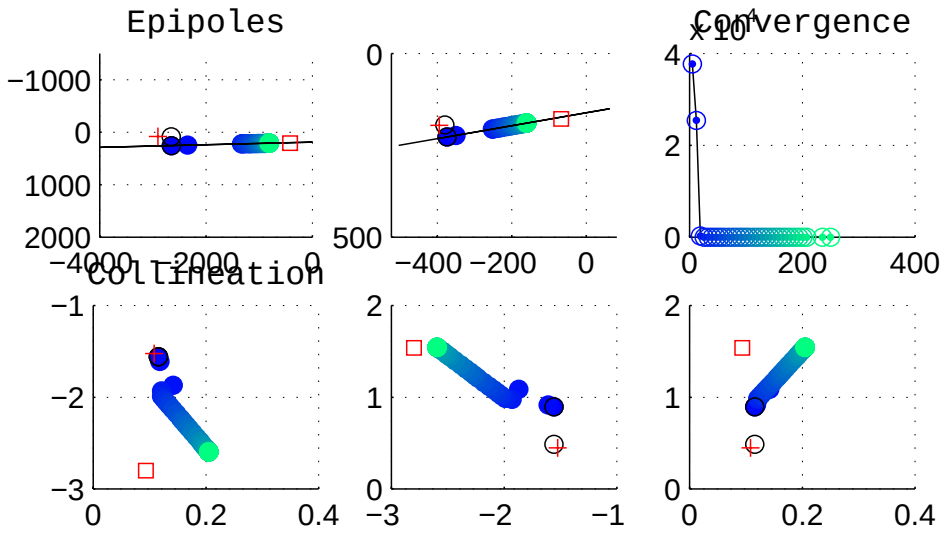


Figure 4-6: *Convergence path for the 4-parameter search for trifocally constrained epipole parameters relating views 1 and 3 of the* `santorini` *sequence. Top left and middle: epipole parameter convergence along the trifocal line. Bottom: Epipolar collineation parameter convergence. Top right: objective function value during search. Solid circles: show the search path, starting at the darkest (blue) and converging at the lightest (green). '+': initial solution found by the 8-point algorithm. Circle: final value found by 7-parameter search. Square: ground truth point.*

Chapter 5

Stable Projective Parameterizations

The 3D analysis of camera geometry, when treated as a parameter estimation problem, requires an evaluation function that measures how well the geometry defined by the parameters fits the observed image data. This chapter examines the properties of the point-to-epipolar-line error function for both the 7D minimal parameterization of a view pair and the 4D trifocally constrained parameterization of two views in a triplet. We examine the shape of the objective function in the neighborhood of its global minimum and find that low-error regions correspond to ambiguities in the projective description of a given Euclidean camera configuration. Empirical studies of synthetic and real imagery illustrate how the statistical uncertainty of trifocal lines influences the trifocally constrained parameterization. Furthermore, we show how the projective parameterizations of nearly degenerate camera configurations lie near unstable points of the parameter space. Finally, we present a stable parameterization of the bifocal and trifocal geometry that avoids such numerical instabilities and improves the reliability of the optimization.

5.1 Objective Functions

The algorithm given in Chapter 4 for fitting a minimal projective model to three views is compatible with any objective function that measures the fit of an epipolar geometry

estimate to a set of matching point pairs. In this work, we use the *point-to-epipolar-line* error function given in Equation 4.1 and restated here for convenience:

$$e_l^2(\mathbf{m}_j, \mathbf{m}_k; \mathbf{p}_{jk}^4, \mathbf{p}_{ij}^7, \mathbf{p}_{ik}^7) \quad = \quad d_l^2(\mathbf{m}_j, \mathbf{l}_j) + d_l^2(\mathbf{m}_k, \mathbf{l}_k). \tag{5.1}$$

To see how this error measures the fit of an epipolar geometry estimate to image point matches, consider the example in Figure 5-1. At any stage in the optimization process, the set of parameters defining the epipolar geometry between two views defines a unique epipole in each view and a mapping from any point in one image plane to an epipolar line through the estimated epipole in the other image plane. Equation 5.1 measures how well a current set of epipolar parameters fits the image matches by summing the squared distances of the observed points to the estimated epipolar lines. This section examines the shape of the evaluation function for various slices of the epipolar parameter space and the trifocally constrained parameter space in order to gain a better understanding of how different estimation methods perform.

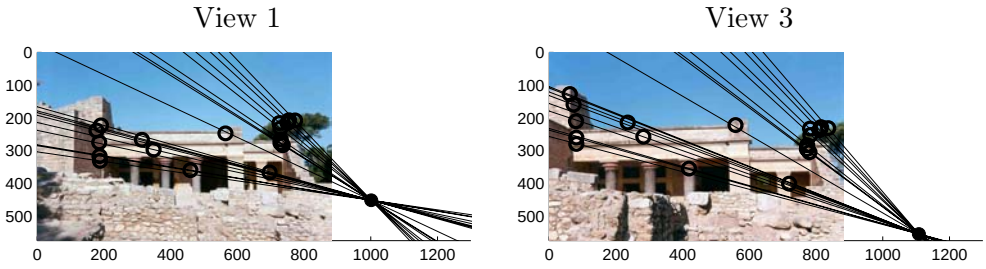


Figure 5-1: *The epipolar lines defined by an estimate of the epipolar geometry between two views. A set of epipolar parameters defines an epipole in each view and a transformation of the pencil of epipolar lines through the points in one image to the epipolar pencil through points in the other image. The point-to-epipolar-line error sums the Euclidean distance of each observed image point to the epipolar line generated from its corresponding image point.*

Although the objective function is theoretically defined for any point in the infinite image plane, in practice, since a captured image has finite image boundaries, the function is only evaluated over points in a small region of the image plane. Consequently, there are potentially many different camera motions that might register similar error on the image

data because they have nearly identical effects on the epipolar lines within the image boundaries. The effect of limited fields of view on uncertainty in camera motion understanding has been studied theoretically and empirically in both the projective and Euclidean frameworks [5, 33, 14].

The point-to-epipolar-line distance is a commonly used evaluation function for the epipolar geometry. In addition, there are a variety of statistically motivated and geometrically meaningful error functions that weigh an observed image point's contribution to the sum by its proximity to the estimated epipole or by the estimated statistical variance of the image point's measured position [53]. These functions may all be used in place of the point-to-epipolar-line distance.

The *true reprojection error* for a camera model is the distance between an observed image point and its predicted position under that model. To use this error for evaluating the epipolar geometry of only two views, the predicted point positions may be added to the set of unknown parameters to estimate. This gives us the cross-epipolar error over the camera parameters and image point observations [29]. If we are willing to estimate the 3D projective positions of scene points by constructing projective camera matrices from the epipolar geometry, then the true reprojection error may also be used, as is done in projective bundle adjustment [34, 3]. While our minimally parameterized framework may be used within both of these frameworks, in this work we choose to perform the optimization only over the camera parameters. The resulting solution may then be used as a starting point for initializing a full-fledged structure and motion optimization in either the uncalibrated projective setting, or after estimating the internal parameters, in the calibrated Euclidean setting.

When three views of a scene point are available, the true reprojection error may be measured without re-estimating image point locations by simply finding the intersection of two epipolar lines in an image [24]. This is equivalent to the point-transfer function used to evaluate how well an estimated trifocal tensor matches a set of matching point triplets [36, 18]. However, since we are dealing with imagery in which many points that are matched across two views are not matched in three views, we choose not to rely on the reprojection error over point triplets. Oliensis [33] and Zhang [54] give comprehensive analyses of the

conditions under which the point-to-epipolar-line error is a good approximation of the true reprojection error.

Let us now examine the shape of the point-to-epipolar-line objective function in the minimal space of epipolar geometry parameters. We compare the error surface in the 7D space of free epipolar parameters to the error surface in the 4D space of trifocally constrained epipolar geometry parameters. In both cases, the error surfaces have elongated valleys that represent regions of uncertainty in the parameter space, as reported in [5].

To visualize the objective function over the epipolar parameters, we show 2D slices of the error value in the neighborhood of the global minimum. Slices from the full 7D space of unconstrained epipolar parameters exhibit shapes similar to slices of the 4D trifocally constrained space of two views in a triplet, except that the global optimum in the constrained space is closer to the true value when the 4D projection forces it to lie in the space of legal solutions.

Consider the Euclidean camera configuration depicted in Figure 5-2. Given a fixed set of internal parameters, the first four views have identical epipolar geometry relative to the fifth camera, which is centered at the origin. This many-to-one relation is due to the four-fold ambiguity in mappings from essential matrices to rotations and translation directions (Section 4.2).

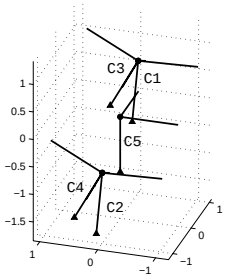


Figure 5-2: *Four Euclidean camera configurations corresponding to the same epipolar geometry relative to the fifth camera.*

The canonical epipolar parameters for this configuration, as defined in Section 2.2.5, are

$$\mathbf{e}_{ij} \simeq [375, 125, 1]^T \tag{5.2}$$

$$\mathbf{e}_{ji} \simeq [125, 125, 1]^T \tag{5.3}$$

$$\mathbf{H}_{ij_{2\times 2}} \simeq \begin{bmatrix} .2576 & -6.31 \\ 6.31 & 1 \end{bmatrix} \tag{5.4}$$

$$\mathbf{p}_{ij}^{7} = \{375, 125, 125, 125, .2576, -6.31, 6.31\}. \tag{5.5}$$

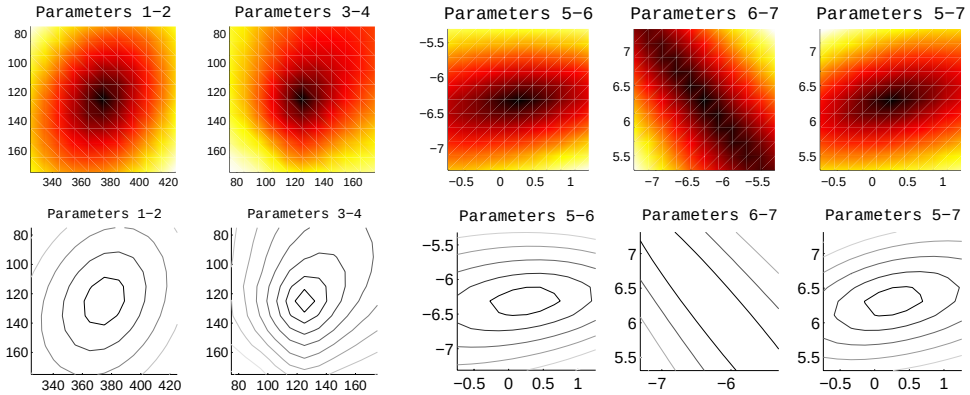


Figure 5-3: *Slices visualizing the objective function over the 7D space of epipolar parameters on noise-free data, centered at the true solution. (a)-(b): Epipole parameter slices. (c)-(e) Collineation parameter slices. Top: Darker regions have lower error values. Bottom: contours of equal error value.*

A block of scene points is generated in front of cameras 1 and 5 and projected into each view. Figure 5-3 shows several 2D slices of the equal error contours of the error function evaluated on the resulting noise-free image point matches. The epipole slices show a well-behaved function with an obvious minimum, as expected when the epipole lies within the image boundaries. The collineation slices show elongated regions of low error. These elongated valleys correspond to the anisotropic covariance ellipsoids known to characterize the uncertainty of the epipole and epipolar collineation parameters [5].

It is apparent that two components of the collineation are strongly anti-correlated: $\mathbf{p}_{ij}^{7}(6) = -\mathbf{p}_{ij}^{7}(7)$. Geometrically, the anti-correlated points along the valley in Figure 5-3(d) correspond to small rotations of the cameras for which the collineation gradually rotates the epipolar line pencil. Figure 5-4 shows the Euclidean interpretations of camera configurations that have the same projective parameters given in Equation 5.5, except for the absolute values of the off-diagonal elements, which are made to vary by a uniform scale factor so

they remain anti-correlated:

$$\mathbf{e}_{ij} \simeq [375, 125, 1]^T \tag{5.6}$$

$$\mathbf{e}_{ji} \simeq [125, 125, 1]^T \tag{5.7}$$

$$\mathbf{H}_{ij\,2\times2} \simeq \begin{bmatrix} .2576 & -n \\ n & 1 \end{bmatrix} \tag{5.8}$$

$$\mathbf{p}_{ij}^7 = \{375, 125, 125, 125, .2576, -n, n\}, \tag{5.9}$$

for $n \in (5, 10, 20, 50)$. Figure 5-4 shows the slight changes of rotation that result from these changes in the collineation parameters. Though the surface in Figure 5-3(d) is shallow, there is a well-defined minimum in this valley that corresponds to the true projective parameters. In practice, we find that as long as the nonlinear optimization enters this valley, it will find the globally optimal solution.

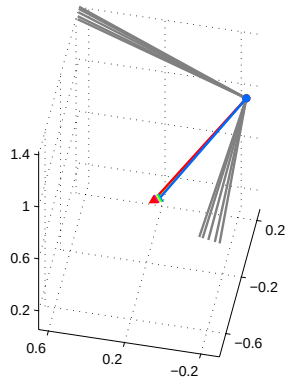


Figure 5-4: *Euclidean projections of multiple points in the valley of the objective function of Figure 5-3(d). The epipolar parameters for these views are identical except for the collineation parameters* $\mathbf{p}_{ij}^7(6)$ *and* $\mathbf{p}_{ij}^7(7)$ *which are uniformly scaled.*

5.2 Statistical Uncertainty of Projective Parameters

Having witnessed the potential ambiguities in projective parameter spaces, let us now examine the effect of image point noise on the error function and how it influences the accuracy of the epipolar geometry estimates. The minimal trifocal model is applicable to situations when the epipolar geometry of two of the camera pairs is accurately estimated before the

remaining constrained epipolar parameters for the third pair are estimated. In practice, image measurement error will result in uncertain estimates of the two initial camera pairs. This section examines the effect of measurement error on the minimal 4-parameter search.

5.2.1 Epipole Covariance

When image point matches are corrupted with additive Gaussian noise, the resulting covariance in epipole estimation using the 8-point algorithm is anisotropic, so the area of uncertainty is an elongated ellipse. Figure 5-5 shows the epipoles derived from multiple trials of the 8-point algorithm on a fixed camera configuration and image point set, with different instances of Gaussian noise added to the image point coordinates for different trials.

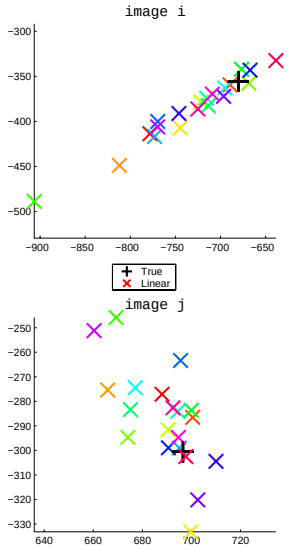


Figure 5-5: *Elongated uncertainty regions for epipoles linearly estimated from repeated trials of Gaussian noise added to image point matches. Each 'x' marks the epipole estimate from a different instance of noisy image data for a fixed camera configuration.*

It is known that the anisotropic uncertainty of epipole estimates is due to the specific geometry of the camera configuration [21]. If a camera translates parallel to the image plane, the epipoles will tend toward infinity and the uncertainty along the direction of translation will grow. The elongated uncertainty region coincides with valleys in the objective function's error surface.

5.2.2 Trifocal Line Error

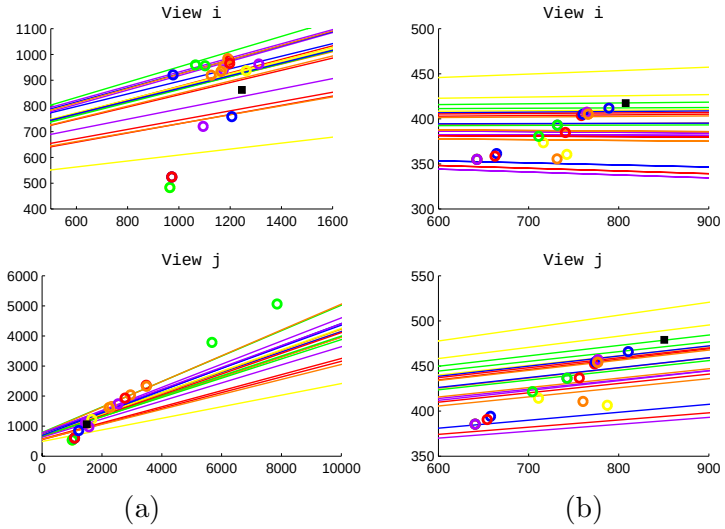


Figure 5-6: *Uncertainty in epipoles vs. trifocal lines. Image points matched between all pairs in three images are corrupted with Gaussian noise. The circles mark independent estimates of the epipoles for one view pair from different instances of noisy image data. The lines mark independent estimates of the trifocal lines estimated using the other two view pairs. (a) Trifocal line estimates properly constrain the solutions when their spread is narrower than the independent epipole estimates. (b) When trifocal line estimates are more uncertain than independent epipole estimates, they do not improve the epipolar geometry estimates.*

Since the minimally constrained trifocal algorithm searches along trifocal lines computed from the uncertain epipoles and collineations relating the two other camera pairs, it is important to understand the error in trifocal line estimate. The following experiments are performed on the simulated 3-view configuration described in an upcoming section (Section 5.5.2). Again, we fix the camera and scene positions and generate 20 instances of noisy image matches. Figure 5-6 shows the distributions of the estimated epipoles and the estimated trifocal lines for each simulation. In each trial, a different subset of image point matches is sampled, using the RANSAC algorithm. When the epipole variance is high and the trifocal line bundle is tightly centered around the epipole, the trifocally constrained model gives a more accurate estimate. When the epipole variance is low enough to lie within the span of the trifocal lines, the unconstrained epipole estimates are more accurate than the constrained ones. Table 5.1 shows the resulting error in Euclidean geometry estimates: for the first example, the trifocally constrained model improves the translation estimate,

	View 1	View 2	View 3		View 1	View 2	View 3
Mean Translation Error : Norm				Mean Rotation Error : Angle			
8−point	6.6782	6.3693	10.9846	8−point	0.0194	0.0206	0.0355
7−parameter	6.3515	4.9028	9.5730	7−parameter	0.0044	0.0053	0.0078
4−parameter	3.3469	3.4576	6.1696	4−parameter	0.0049	0.0054	0.0091

(a)

	View 1	View 2	View 3		View 1	View 2	View 3
Mean Translation Error : Norm				Mean Rotation Error : Angle			
8−point	2.4959	2.8364	3.9781	8−point	0.0098	0.0124	0.0221
7−parameter	2.0947	2.2879	3.2396	7−parameter	0.0072	0.0078	0.0147
4−parameter	1.8243	2.2509	2.9978	4−parameter	0.0052	0.0058	0.0107

(b)

Table 5.1: *Mean errors in camera translations and rotations for the two examples in Figure 5-6. (a) The minimal, trifocally constrained model improves the translation of the unconstrained estimate on average when the trifocal line variation is lower than the epipole variation. (b) The minimal trifocal model offers only a slight improvement when the trifocal lines vary more than the epipoles.*

while in the second example it is slightly worse than the unconstrained epipolar geometry estimate.

5.3 Critical Configurations

The regions of uncertainty shown in the Section 5.1 do not pose a problem for optimization procedures assuming a reasonable initial point. However, when the initial point of the search sits at a critical point of the objective function, there is a danger of entering and becoming trapped in a valley that contains a local minimum. This section examines camera configurations whose canonical parameterization does not lie at a stable point. Section 5.4 presents a generalized parameterization of the epipolar geometry that avoids such instabilities.

View Pairs

The two-view models used in this work are for general motions, meaning they allow all possible 3D translations, 3D rotations, and variations in all 5 internal parameters. When

a camera motion is not rich enough to use all of these degrees of freedom, the epipolar geometry takes on special forms. We consider the two common cases of pure translation and coplanar images, and their effects on the epipolar geometry. By understanding these cases, we can adapt the canonical parameterization to be numerically stable even when the camera motion is near a critical configuration.

In the case of pure translation, the fundamental matrix takes the form,

$$\mathbf{F}_{ij} \simeq \begin{bmatrix} 0 & - & - \\ - & 0 & - \\ - & - & - \end{bmatrix}, \tag{5.10}$$

where the components marked by '-' depend on the camera translation and the internal parameters. If we decompose the fundamental matrix in the canonical way, so the epipolar collineation is defined by the upper left submatrix of the fundamental matrix (Section 2.2.5), the collineation takes the form,

$$\mathbf{H}_{ij\,2\times2} \simeq \begin{bmatrix} 0 & - \\ - & 0 \end{bmatrix}.$$

Clearly, the component of $\mathbf{H}_{ij\,2\times2}$ chosen for normalization (Section 2.2.4) should not be small, so for small rotations it is important to choose this component with care to avoid division by small values.

In the case where two cameras have parallel image planes and the 3D camera motion is pure translation, the fundamental matrix takes the form,

$$\mathbf{F}_{ij} \simeq \begin{bmatrix} 0 & 0 & - \\ 0 & 0 & - \\ - & - & - \end{bmatrix}. \tag{5.11}$$

This is known as the affine fundamental matrix [35], and its epipoles are at infinity, so they take the form,

$$\mathbf{e}_{ij} \simeq [-, -, 0]^T$$

$$\mathbf{e}_{ji} \simeq [-, -, 0]^T,$$

as can be verified by checking that $\mathbf{F}_{ij}\mathbf{e}_{ij} = \mathbf{0}$ and $\mathbf{F}_{ij}^T\mathbf{e}_{ji} = \mathbf{0}$. Only the first or second components of the epipoles may therefore be used for normalization. If the translation is horizontal, the epipoles are still at infinity, and they now take the even simpler form,

$$\mathbf{e}_{ij} \simeq [-, 0, 0]^T$$
$$\mathbf{e}_{ji} \simeq [-, 0, 0]^T.$$

Figure 5-7 shows shows this scenario. Once again, it is important to choose a nonzero normalization component for each epipole to ensure numerical stability.

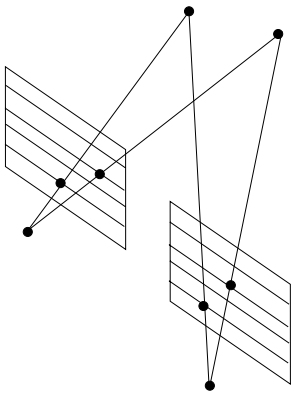


Figure 5-7: *The affine epipolar geometry. The 3D camera translation is parallel to both image planes, so the epipoles lie at infinity and the epipolar lines are horizontal.*

The problem of multiple model selection for camera geometry is an important topic of study [44, 23]. While a truly degenerate camera motion calls for a simplified model, in practice there is a continuum of increasingly simpler camera motions, and it is unclear at what point a new model should be adopted. It is nevertheless useful to examine the effects of these motions on the shape of the objective function and the pitfalls they pose for nonlinear estimation. This analysis leads us to generalize the existing canonical parameterization to handle near-degeneracy without losing valuable information by fitting the data to a a simplified model.

The next section examines what the objective function looks like for various synthetic

critical camera motions. Section 5.4 introduces a generalization of the canonical parameterization that takes these potential singularities into consideration. Section 5.5.2 presents experimental results of applying this robust parameterization to the trifocally constrained parameter estimation algorithm from the previous chapter.

Error Surfaces

Consider the camera configuration shown in Figure 5-8. The motion between the two cameras is a nearly pure translation, and the camera planes are nearly parallel.

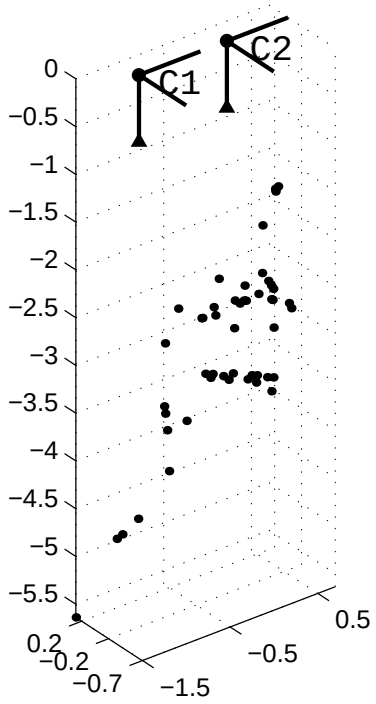


Figure 5-8: *Synthetic critical camera configuration.*

Now let us introduce internal parameters, construct a projective parameterization of this configuration, and examine the objective function measured over the imaged scene points. Figure 5-9 shows the viewed image points for a focal length of 500 pixels, no skew in the camera axes, and principal point centered in the image. The canonical projective parameterization is

$$\mathbf{e}_{ij} \simeq [100000, 10, 1]^T$$

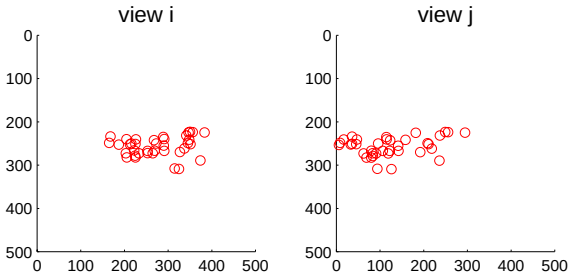


Figure 5-9: *Visible image points for the synthetic configuration in Figure 5-8, using a focal length of 500 pixels.*

$$\begin{aligned} \mathbf{e}_{ji} &\simeq [100000, 10, 1]^T \\ \mathbf{H}_{2\times 2_{ij}} &\simeq \begin{bmatrix} 0 & -14704 \\ 14704 & 1.0 \end{bmatrix} \\ \mathbf{p}_{ij}^7 &= \{100000, 10, 100000, 10, 0, -14704, 14704\}. \end{aligned}$$

Figure 5-10 shows the level contours of the objective function. The elongated regions of low error are due to the expected ambiguity in the solution: the y-coordinates of the epipoles are well-constrained by the image data, but the x-coordinates are uncertain because they have large horizontal components and small vertical components. The horizontal epipolar lines are largely unaffected by local changes in the x-coordinates of the epipoles. The upper left component of the epipolar collineation, $\mathbf{p}_{ij}^7(5)$, is not sensitive to small changes, but the two off-diagonal components, $\mathbf{p}_{ij}^7(6)$ and $\mathbf{p}_{ij}^7(7)$, are strongly anti-correlated. The ratios of the collineation parameters are more important than their absolute values. The off-diagonal elements may vary uniformly as long as their magnitudes are much larger than the diagonal elements, and the error will still be low because the epipolar geometry will remain nearly affine. In practice, we find that if this valley is entered by the nonlinear optimization, a global minimum which corresponds to the optimal solution will be found.

Now examine the objective function in the neighborhood of a different point in the same parameter space by changing the collineation parameters as follows:

$$\begin{aligned} \mathbf{e}_{ij} &\simeq [100000, 10, 1]^T \\ \mathbf{e}_{ji} &\simeq [100000, 10, 1]^T \end{aligned}$$

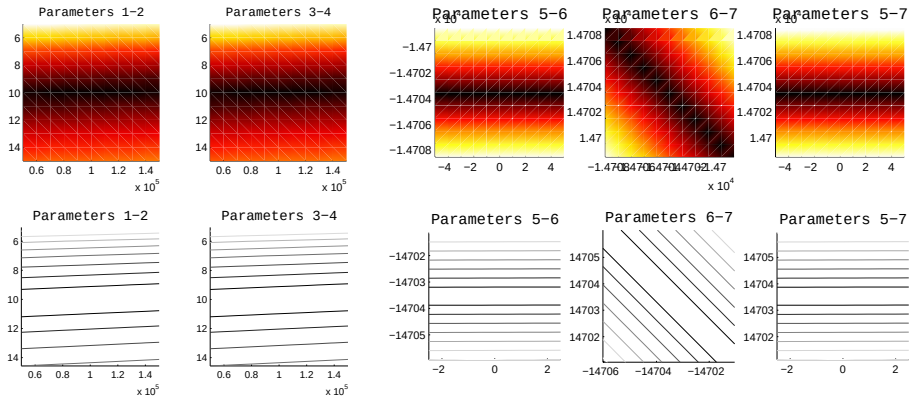


Figure 5-10: *Slices visualizing the objective function over the 7D space of epipolar parameters on noise-free data for a nearly affine epipolar geometry. See text for explanation. (a)-(b): Epipole parameter slices. (c)-(e) Collineation parameter slices. Top: Darker regions have lower error values. Bottom: contours of equal error value.*

.

$$\begin{aligned} \mathbf{H}_{2\times 2_{ij}} &\simeq \begin{bmatrix} 0 & -1.0562 \\ 1.061 & 1.0 \end{bmatrix} \\ \mathbf{p}_{ij}^{7} &= \{100000, 10, 100000, 10, 0, -1.0562, 1.061\}. \end{aligned}$$

This point in parameter space corresponds to the same Euclidean configuration but a different set of internal camera parameters. We see in Figure 5-11 that there is a saddle point leading to different valleys separated by high-error peaks because the ratios of the collineation elements are unstable near the origin. If the nonlinear optimization is initialized in this region of the search space, there is a danger that it will fall into the wrong valley and become trapped in a local minimum. This synthetic example highlights the importance of choosing a parameterization for which the optimal solution is not near an unstable point of the parameter space. The next section presents a generalization of the canonical parameterization which automatically takes into account the geometry of the configuration to avoid such coordinate singularities and numerical instabilities. Sections 5.5.1 and 5.5.2 present empirical results for these scenarios.

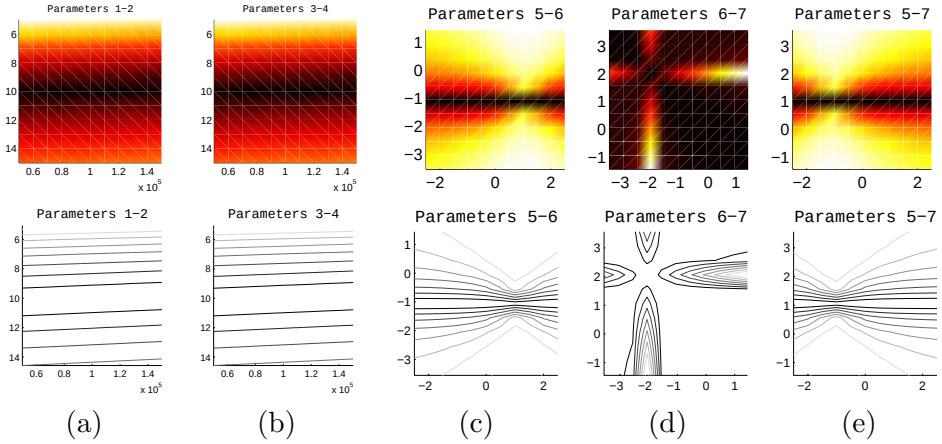


Figure 5-11: *When the solution lies near a saddle point of the objective function, there is a higher chance that the search algorithm will be initialized close to a valley that contains a local minimum instead of the global minimum. (a)-(b): Epipole parameter slices. (c)-(e) Collineation parameter slices. Top: Darker regions have lower error values. Bottom: contours of equal error value.*

View Triplets

In the case that the critical configurations discussed above occur for pairs of views in an image triplet, Section 5.4 shows how to ensure that the trifocally constrained epipolar parameter model is still well-defined, as long as the trifocal plane among three views is well-defined.

The only configuration in which the trifocal plane is not defined is when the optical centers of the three cameras are collinear. When the optical centers are aligned, the trifocal plane does not exist, so the trifocal lines are not defined (Section 2.3.4). This scenario may be detected at the initialization stage of the estimation algorithm in Section 4.1 by noting that the epipolar geometries of one view with respect to each of the other two views are identical or nearly identical. In Figure 2-10, since the baseline directions among all three views are identical, the epipoles for view pair (i, j) are identical to the epipoles for the other view pairs: $\mathbf{e}_{ij} \simeq \mathbf{e}_{ik}$ and $\mathbf{e}_{ji} \simeq \mathbf{e}_{ki}$. Furthermore, the epipolar planes are identical for all pairs, so the epipolar collineation for pair (i, j) is simply a composition of the two other collineations: $h_{ij}(\mathbf{l}_i) \simeq h_{kj}(h_{ik}(\mathbf{l}_i))$. Therefore, once the epipolar geometries of two of the pairs is fixed, the epipolar geometry of the third pair is completely determined.

In practice, one might examine the three initial estimates of the pairwise epipolar geometries, and use the similarities of the epipoles and collineations to decide whether or not the views are collinear and hence whether or not to apply the trifocal constraints to the epipolar parameters. The general problem of detecting model degeneracy is an important issue in multiple view camera analysis and is discussed briefly in Chapter 6.

In order to obtain a complete projective representation of a collinear camera set, it is necessary to use the trilinear constraints on point triplets [36, 13]. Because of these scenarios, the problems of degeneracy detection and multiple model hypothesis selection are important issues in projective camera modeling. Torr presents a case-based algorithm for using information criteria to choose the simplest projective model that is not degenerate for a given set of data [44].

5.4 Stable Parameterization

This section generalizes the canonical parameterization of the epipolar geometry to improve its numerically stability during estimation. Recall that the canonical parameterizations given in Section 2.2.4 are not unique. First, there are $3 \times 3 \times 4$ combinations of normalization indices for the epipoles and the epipolar collineation, each resulting in a distinct minimal representation of the same epipolar geometry. In theory, these parameterizations are all equivalent in that they produce the same epipolar pencils and therefore enforce the same constraints on matching image points. In practice, however, they differ in numerical stability, causing widely differing behaviors during optimization. Statistical methods for finding the choice of normalization components that results in a numerically well-behaved decomposition of an estimated fundamental matrix have been explored in [5]. However, this countable set of parameterizations only corresponds to those minimal parameterizations that use canonical basis elements to represent the epipolar pencils (Equations 2.32 to 2.35).

This section presents a more general class of parameterizations of the epipolar geometry for two views and for three trifocally constrained views. The parameterization is designed to be stable and well-defined in the parts of the search space explored during optimization.

It allows an arbitrary choice of basis for the epipolar pencil, and an arbitrary choice of basis points for the trifocal line. The 7-parameter and 4-parameter parameterizations are automatically chosen, based on the initial estimate of the epipolar geometry. Sections 5.5.1 and 5.5.2 give experimental evidence that the extended parameterization results in accurate estimates of the epipolar geometry.

5.4.1 Normalization

The canonical minimal parameterization of the epipolar geometry given in Section 2.2.4 requires a choice of indices for normalizing the epipole and the epipolar collineation. Clearly, these components should be chosen carefully to avoid division by very small values. Csurka, *et al.*, address this issue by choosing the normalization indices that maximize the rank of the Jacobian of the mapping from fundamental matrices to the epipolar parameters at each step of the optimization [5]. For the epipole normalization, we adopt a simpler method of choosing the largest component of the initial epipole estimate, an approach also used in [2]. Typically, in real imagery with primarily horizontal translation, this will be the x-component, so the resulting epipole representation takes the form $[1, -, -]^T$. The same step is taken for the epipolar collineation. To avoid division by a small value, the largest component is chosen for normalization. The next section presents a more general method of representing the epipolar collineation that does not depend solely on the epipole normalization index for stability.

For the trifocally constrained epipolar parameterization, recall that one component of the epipolar collineation is chosen to be dependent on the other collineation components (Equation 3.29). To ensure that the expression is well-defined, the choice of index c_3 for the dependent collineation component must be made such that the coefficient $\mathbf{k}(c_3)$ is not close to zero. From Equation 3.27, it is clear that if we can guarantee that the representations $[\alpha_j, \beta_j]^T$ and $[\alpha_k, \beta_k]^T$ of the trifocal lines have nonzero components, then the coefficients $\mathbf{k}$ will be nonzero. The next section provides a geometrically principled way of automatically finding such a parameterization.

5.4.2 Epipolar Pencil Basis

Recall from Sections 2.2.4 and 2.2.5 that the parameterization of the epipolar collineation depended on the basis points $(\mathbf{p}_j, \mathbf{q}_j)$ and $(\mathbf{p}_k, \mathbf{q}_k)$ used to represent the epipolar pencils in views j and k, respectively (Figure 5-12). Choosing these to be canonical basis vectors allowed the collineation elements to be directly observable as components of the fundamental matrix. However, these basis points are not necessarily good representations of the epipolar pencil. For example, if the epipoles are represented by their first two coordinates, the canonical canonical basis vectors are the horizontal and vertical points at infinity. If the epipoles lie to the left or the right of the image, then this representation will have a high coefficient for one basis point and a low coefficient for the other basis point (Figure 5-13).

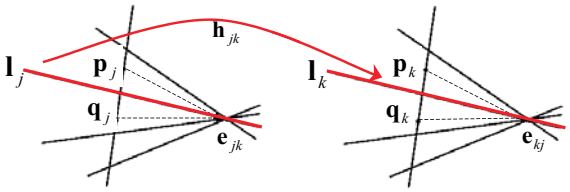


Figure 5-12: *Two basis points in each view define the epipolar pencil as a 1-parameter family.*

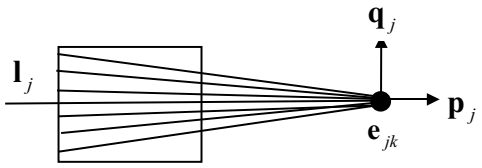


Figure 5-13: *An example of when the canonical basis vectors are a poor representation of the epipolar pencil. The epipolar lines for many observed image points will have a large component for* $\mathbf{p}_i$ *and a vanishing component for* $\mathbf{q}_i$.

The generalized parameterization chooses the basis points that ensure a balanced set of coefficients for the epipolar lines that pass through the image. Consider Figure 5-14, and suppose the labeled epipole is an initial estimate or an intermediate estimate found during optimization. Then the two basis points for parameterizing the pencil may be chosen to lie equidistant from the image center along a line that is normal to the line containing the image center and the estimated epipole. The distance of each basis point is chosen to be of

the same order of magnitude as the epipole estimate, measured in image pixels. Using this choice of basis points, we guarantee that the two coefficients in Equations 2.15 and 2.20 are equal and nonzero.

Note that these are precisely the desired conditions for the coefficients of the trifocal line representation to be stable, as discussed in the previous section on normalization, since it guarantees that the denominator of Equation 3.29 is nonzero. However, since the trifocal line does not necessarily cross through the interior of the image, there may be a trade-off in how well the epipolar pencil basis represents the epipolar lines through the observed data points vs. how well it represents the trifocal line in specific.

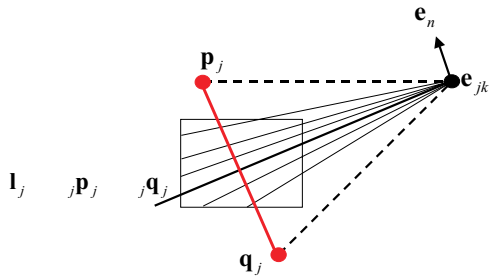


Figure 5-14: *Generalized parameterization of the epipolar geometry. The basis points of the epipolar pencil are automatically chosen to lie equidistant from the image center along the normal* $\mathbf{e}_n$ *to the estimated epipole* $\mathbf{e}_{ij}$*. The coefficients* α_i *and* β_i *representing the epipolar lines through the image are well-balanced and nonzero.*

5.4.3 Trifocal Line Basis

In the trifocally constrained model, each epipole is represented by a single parameter along the trifocal line (Section 3.2). The choice of points $(\mathbf{x}_j, \mathbf{y}_j)$ on the trifocal line $\mathbf{t}_j$ (Equation 3.19) affects the numerical stability of the epipole search since the parameter γ_j will tend toward 0 or infinity as the epipole approaches the basis points. To automatically compute a stable parameterization of the trifocal line, the two basis points are chosen to be equidistant from the estimated epipole and of the same order of magnitude, as depicted in Figure 5-15.

The next two sections illustrate the stability and accuracy of estimating the generalized parameterization.

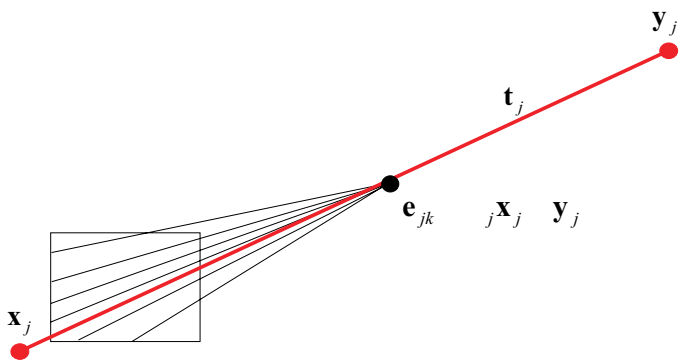


Figure 5-15: *A stable basis for the trifocal line is chosen by finding points equidistant from the epipole and of the same order of magnitude as the epipole norm, measured in image pixels.*

5.5 Experimental Results

5.5.1 Synthetic Sequence

To test the estimation of the trifocally constrained epipolar geometry using the generalized parameterization from the previous section, we construct a set of cameras for which view pair (1,3) has a nearly affine epipolar geometry. Figure 5-16 shows the ground truth Euclidean camera poses and the synthetic image points generated using a focal length of 500 pixels, orthogonal image axes, and image-centered principal point.

The image points are corrupted by Gaussian noise with $\sigma = .5$. The canonical parameterizations of the initial linearly estimated epipolar geometry are highly unstable and the estimation is quickly trapped in a local minimum far from the true epipolar geometry. Figure 5-17 shows the error surface centered at the true solution when the epipoles are normalized by their first components. Similar saddle points are found in the 4-parameter space.

Figure 5-18 shows the error surface near the true epipolar geometry for the generalized parameterization, where now the basis points of the epipolar pencil are automatically chosen based on the initial linear estimate of the epipolar geometry. The error surface still has elongated regions of uncertainty, but the saddle points are no longer in the vicinity of the point representing the true solution. The estimated parameters are shown in Figure 5-19, and the percent errors of the epipole estimates in Figure 5-20. The trifocally constrained

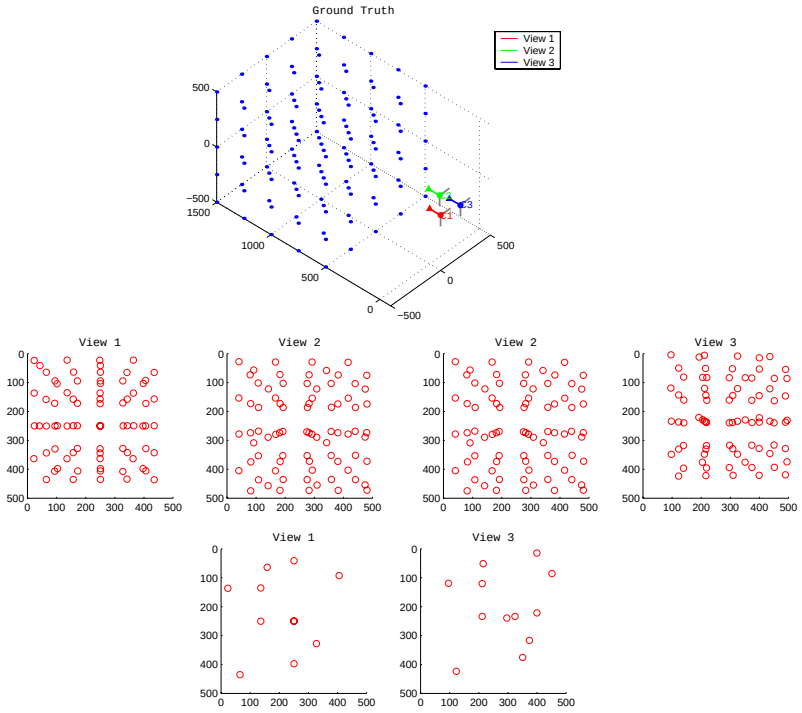


Figure 5-16: *Synthetic camera configuration and image point matches for testing the estimation of epipolar geometry using the generalized parameterization. The point matches between view pair (1,3) are decimated to simulate the asymmetric availability of matches among pairs in a triplet.*

solution improves the accuracy of the epipolar parameters. Using the synthetic internal parameters to project the projective representation to a Euclidean camera configuration, we find that the trifocal constraints improve the accuracy of the camera poses as well (Figure 5-21). The errors in translations and rotations are given in Table 5.2.

5.5.2 Real Image Sequence

This section analyzes an image sequence for which the relative motion between each camera pair is nearly pure translation. The example illustrates how the canonical parameterization of the true camera configuration lies close to a discontinuity in the objective function over the epipolar parameters. In addition to highlighting the importance of choosing a numerically well-behaved projective parameterization, the example reveals how the valleys of the objective function correspond to families of epipolar parameters that correspond to nearly

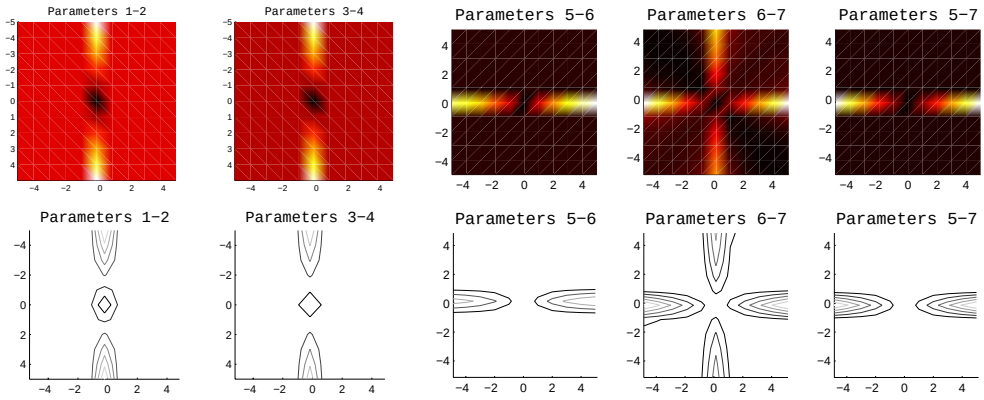


Figure 5-17: *Saddle points in the objective function surface for the unstable canonical parameterization of the synthetic sequence.*

identical Euclidean configurations.

The same procedure is used to set up the simulation as was used for the `santorini` sequence in Section 4.3. The ground truth is constructed by first estimating the camera geometry from matching point pairs found by the most lenient mode of Zhang's image-matching algorithm [55], with false matches manually removed. Figure 5-22 shows the three pairs of frames and their matched feature points. The synthetic ground truth for camera placement and scene structure is defined by independently estimating the three pairwise epipolar geometries and projecting them to a Euclidean configuration (Section 4.2). The focal length is defined to be 1500 pixels, the optical axis is defined to intersect the image center, and the image axes are defined to be orthogonal. Scene points are triangulated, and then projected into each view. Only the reprojected images of the visible points from the original image triplet are used as true, noise-free image point matches. Figure 5-23 shows the simulated ground truth camera configuration, scaled, translated, and rotated to a realistic world coordinate frame.

To compare methods, Gaussian noise with unit pixel variance is added to the synthetic image points. Figure 5-24 compares the linear 8-point algorithm, the independent nonlinear 7-parameter optimization, and the trifocally constrained nonlinear 4-parameter optimization of the epipolar geometry between views 1 and 3. The epipole estimation is greatly improved by the minimally parameterized search, but the collineation estimate is incor-

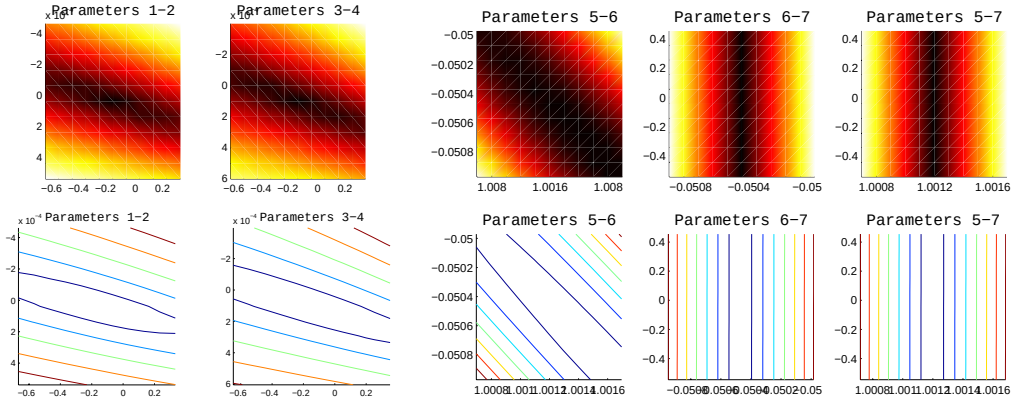


Figure 5-18: *Objective function surface and equal error contours for the generalized parameterization of the synthetic sequence. The region surrounding the true solution does not have saddle points, so the minimum is easily found, as shown in Figure 5-19.*

rect. This behavior is due to the numerical instability of the canonical parameterization, as discussed in the previous section.

The shape of the error surface for the canonical parameterization is shown in Figure 5-25. The collineation slices of the 7D space show that the true solution lies near a saddle point. The 4D trifocally constrained projection lands in the wrong valley, as can be seen in the collineation slice of Figure 5-26.

The convergence paths for these optimizations show that the trifocal line estimate provides a good constraint for the epipole search (Figure 5-27). However, the collineation search is trapped in a valley whose local minimum is far from the true solution. As discussed in Sections 5.3, the origin is often an unstable point for the collineation parameterization. When the solution lies far from this critical point, there is little danger of searches falling into local minima, but when the initial estimate lies close to the discontinuity, the projection onto the trifocally-constrained surface can trap the solution in the wrong valley, as in this case. To avoid this behavior, it is essential to choose a stable parameterization.

The generalized parameterization of the `knossos` sequence is automatically computed from the initial linear estimate of the epipolar geometry. Figure 5-28 shows the pencil basis points and trifocal line basis points chosen for views 1 and 3 of the sequence.

Figure 5-29 shows the solutions for the three different algorithms using the generalized

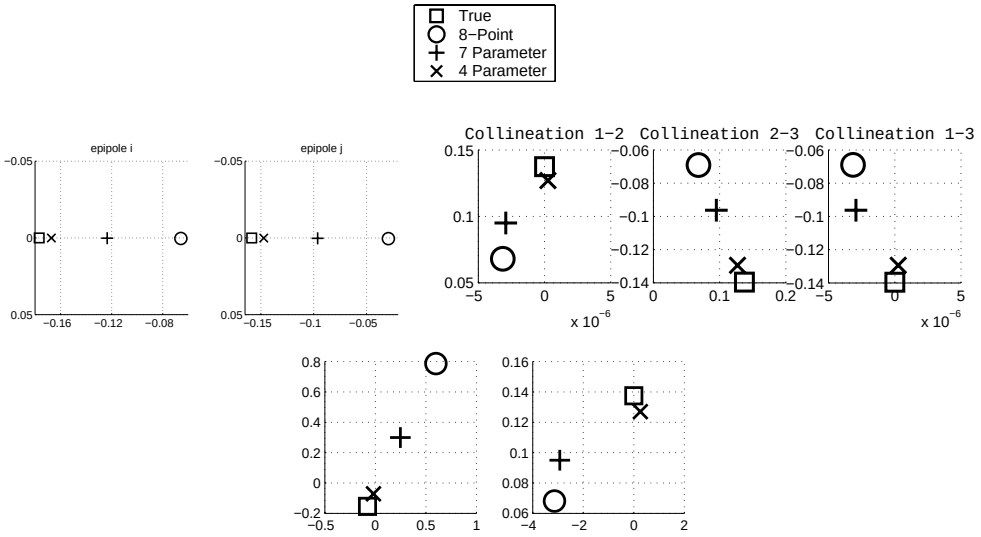


Figure 5-19: *Comparison of epipolar parameters found by three methods on noisy image points for synthetic image sequence. The trifocally constrained epipolar geometry estimate, marked by an 'x', is the most accurate. The square marks the true solution. The errors in the collineation parameters are small due to the small amount of rotation between views.*

parameterization. Figures 5-30 and 5-31 show that the objective functions for the trifocally constrained estimation are no longer near saddle points. Figure 5-32 shows the percent error in epipole estimate for the three methods. Figure 5-33 shows the Euclidean projections of the epipolar parameters, and Table 5.3 gives the errors.

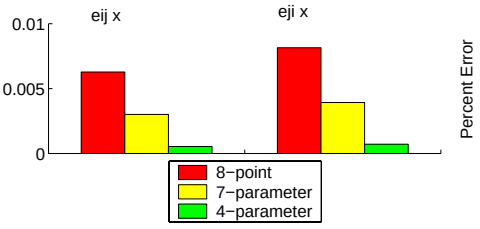


Figure 5-20: *Percent error in x-coordinate of epipole estimates for three compared methods. The error in the estimation of the epipole y-coordinates is negligible for all three estimates because the translation between views is horizontal.*

	3D Translation Error : Norm		3D Translation Error : Angle (Radians)		3D Rotation Error : Angle (Radians)	
	View 1	View 3	View 1	View 3	View 1	View 3
$8-$ Point	32.0306	24.7442	0.0004	0.0315	0.2963	0.0923
$7-$ Parameter	13.8980	10.5767	0.0007	0.0155	0.1104	0.0429
$4-$ Parameter	1.6989	2.9218	0.0004	0.0032	0.0113	0.0129
	Translation Error : X		Translation Error : Y		Translation Error : Z	
	View 1	View 3	View 1	View 3	View 1	View 3
$8-$ Point	20.7764	5.1052	21.1159	20.1702	12.1826	13.3931
$7-$ Parameter	8.6674	1.1581	9.6358	8.4125	5.0182	6.3052
$4-$ Parameter	0.8525	1.3021	1.2774	2.5065	0.7267	0.7478

Table 5.2: *Errors in 3D camera translations and rotations for synthetic sequence, projected from the epipolar geometry estimates to Euclidean parameters using the synthetic internal camera parameters. Each estimate's camera triplet is registered to the ground truth by alignment of the middle view.*

	3D Translation Error : Norm		3D Translation Error : Angle (Radians)		3D Rotation Error : Angle (Radians)	
	View 2	View 3	View 2	View 3	View 2	View 3
$8-$ Point	4.6415	27.9907	0.0439	0.1857	0.0030	0.1030
$7-$ Parameter	4.6805	25.0232	0.0322	0.1742	0.0054	0.0939
$4-$ Parameter	4.6805	8.7066	0.0322	0.0698	0.0054	0.0336

Table 5.3: *Errors in 3D camera translations and rotations for generalized parameterization of the* `knossos` *sequence, projected from epipolar geometry estimates to Euclidean parameters using the synthetic internal camera parameters. Each estimate's camera triplet is registered to the ground truth by alignment of the first view.*

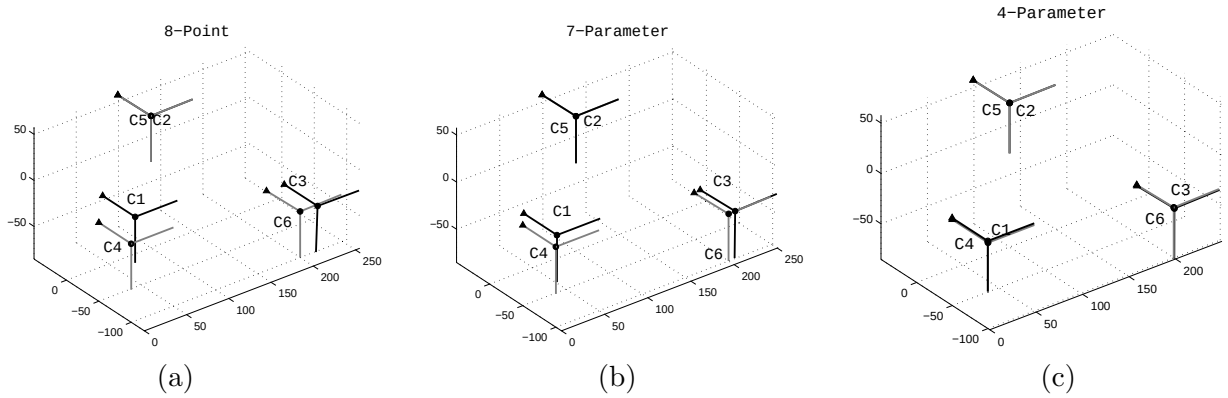


Figure 5-21: *Camera poses found by projecting the epipolar geometry estimates for the synthetic sequence to Euclidean rotations and translations. Estimates are superimposed over ground truth. Camera labels (1,2,3) are the estimates and camera labels (5,6,7) are the corresponding ground truth poses. Triplets are aligned with view 2 and plots are in world units. (a) Linear 8-Point algorithm. (b) Nonlinear 7-Parameter optimization without trifocal constraints. (c) Nonlinear 4-Parameter optimization of trifocally constrained model.*

Figure 5-22: *Point correspondences for* `knossos` *sequence. View pair (1,2) has 60 matches; pair (2,3) has 38 matches; pair (1,3) has 18.*

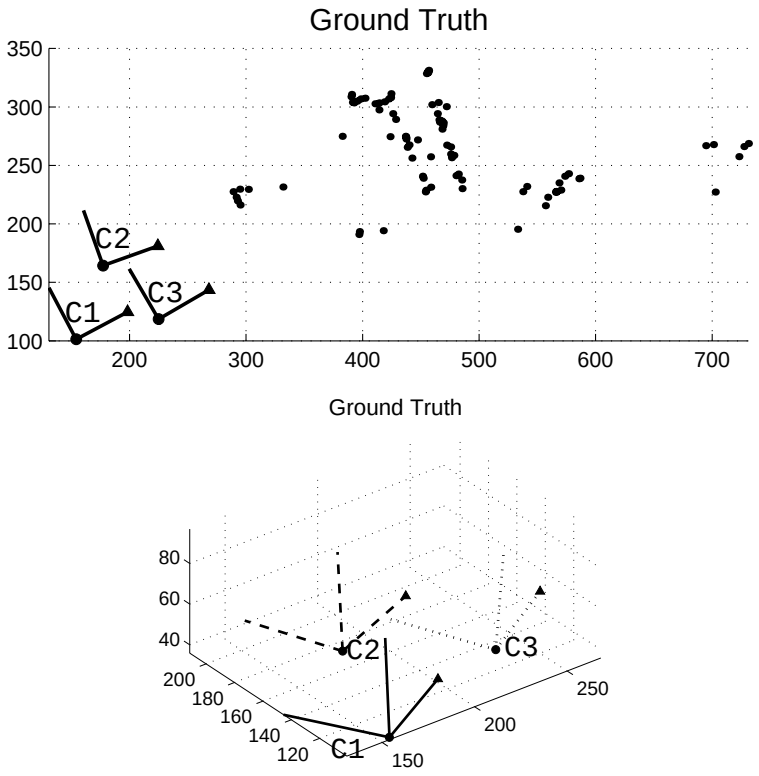


Figure 5-23: *Synthetic ground truth for* `knossos` *image sequence simulation. (a) Overhead view of cameras and reconstructed scene points. (b) Side view of camera poses.*

.

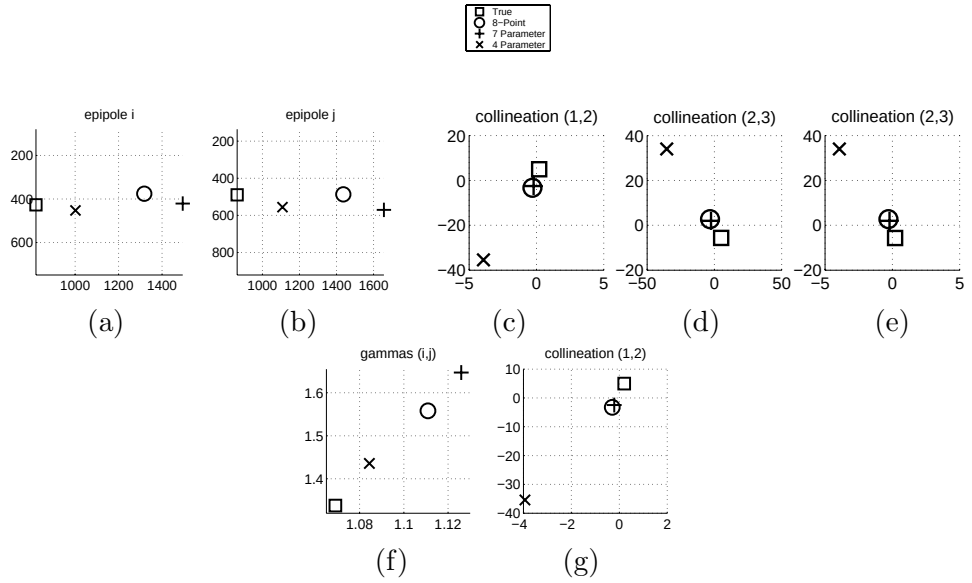


Figure 5-24: *Estimates of the epipolar geometry parameters for views 1 and 3 of the* `knossos` *sequence simulation. (a-b) Components of the reconstructed epipoles, embedded in Euclidean space and given in pixel coordinates. (c-e) Three free components of the reconstructed collineation matrix defined up to scale. (f-g) Trifocally constrained projection of each estimate. (e) Two epipole parameters along the trifocal line; (f) Two free collineation parameters. The epipole accuracy is improved by the minimally constrained model, but the collineation diverges. Compare to Figure 5-29.*

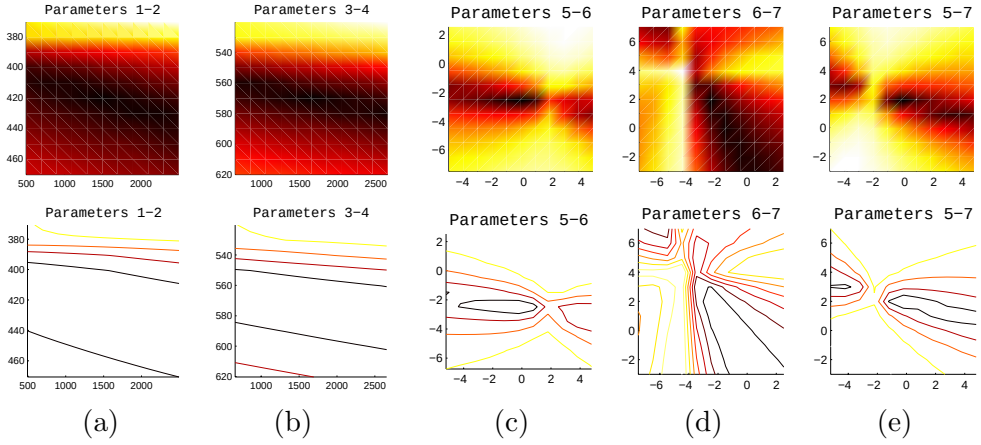


Figure 5-25: *Top: Objective function for the 7D bifocal parameter space on noisy data centered at the 7-parameter estimate. Bottom: Equal error contours. (a-b) Epipole parameter slices. (c-e) Collineation parameter slices. The optimal solution lies near a saddle point of the objective function.*

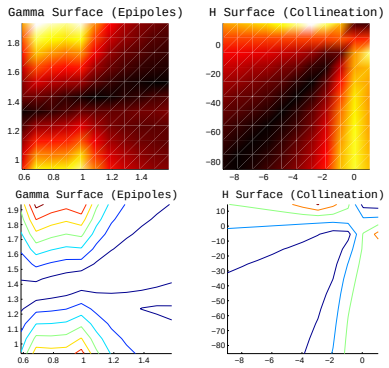


Figure 5-26: *Top: Objective function for* `knossos` *simulation. Slices of error in minimal 4D parameter space on noisy point matches, centered at the 4-parameter estimate. Bottom: Equal error contours. (a) Epipole parameter slices. (b) Collineation slices. The projection onto the trifocally constrained parameter space sends the 4-parameter estimate into the wrong valley of the objective function.*

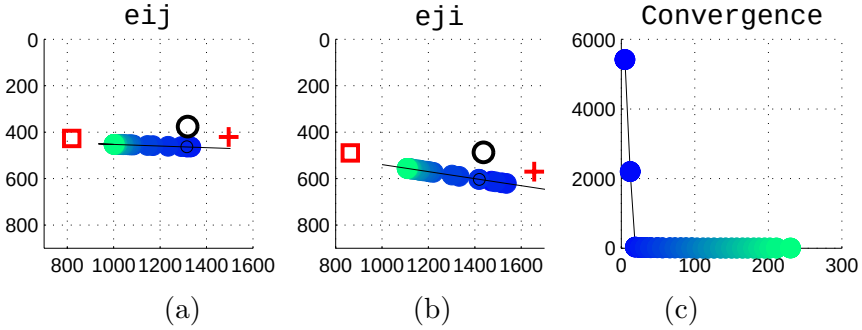
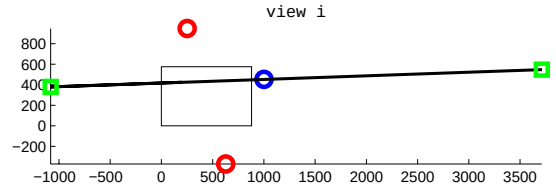
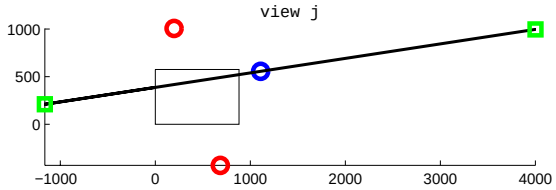


Figure 5-27: *Convergence of the 4-parameter search for* `knossos` *simulation. (a-b) Epipole parameter convergence for 4-parameter search along the trifocal line. (c) Objective function value during search. Solid circles show the search path, starting at the darkest (blue) and converging to the lightest (green). The initial solution found by the 8-point algorithm marked by a '+'. The 7-parameter solution is marked by an open circle. The ground truth is marked by a square.*



(a)



(b)

Figure 5-28: *Basis for the stable parameterization, computed from initial epipole estimates. The rectangle indicates the image boundaries; the epipole is marked by the blue circle to the right of the image; the basis points for the epipolar pencil are marked as red circles above and below the image; the basis points for the trifocal line are marked as green squares.*

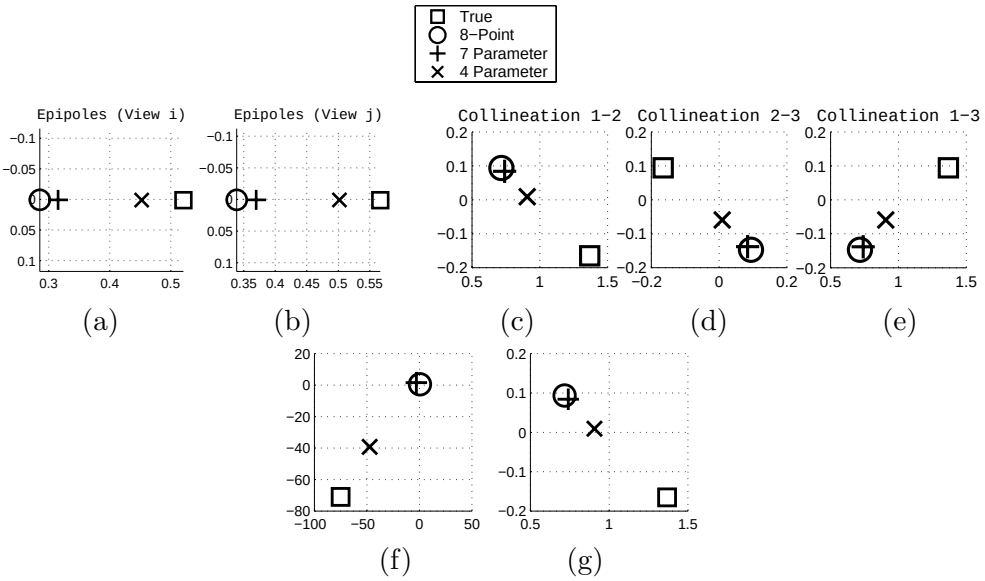


Figure 5-29: *Estimates of the epipolar geometry parameters for views 1 and 3 of the* `knossos` *sequence simulation using the generalized parameterization. (a-b) Components of the reconstructed epipoles, embedded in Euclidean space and given in pixel coordinates. (c-e) Three free components of the reconstructed collineation matrix defined up to scale. (f-g) Trifocally constrained projection of each estimate. (e) Two epipole parameters along the trifocal line; (f) Two free collineation parameters. Accuracy of all parameters is improved by the trifocally constrained model. Compare to Figure 5-24.*

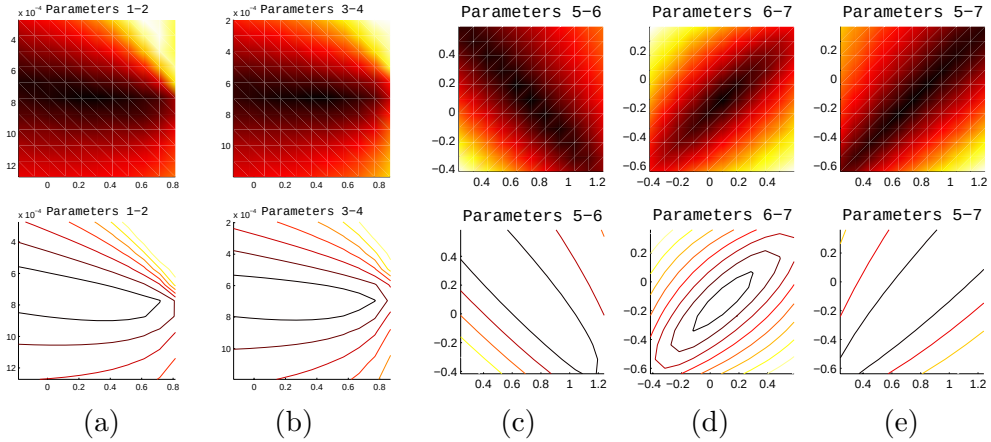


Figure 5-30: *Top: Objective function for the generalized parameterization of the* `knossos` *simulation. Slices of the error in 7D space, on noisy point matches. Darker shade indicates lower error. Bottom: equal error contours (a-b) Epipole parameter slices. (c-e) Collineation slices.*

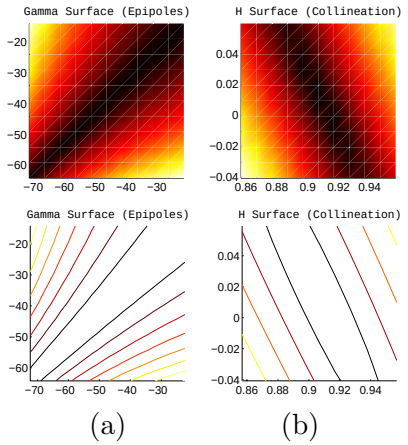


Figure 5-31: *Top: Objective function for the generalized parameterization of the* `knossos` *simulation. Slices of error in minimal 4D parameter space on noisy point matches, centered at the 4-parameter estimate. Darker shade indicates lower error. Bottom: Equal-error contours. (a) Epipole parameter slices. (b) Collineation slices.*

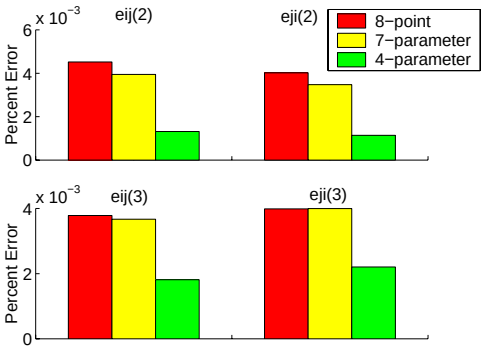


Figure 5-32: *Percent error in epipole estimates for the three compared methods using generalized parameterizations. Epipoles are normalized by their first coordinates. Top: first epipole parameter for both views. Bottom: second epipole parameter for both views.*

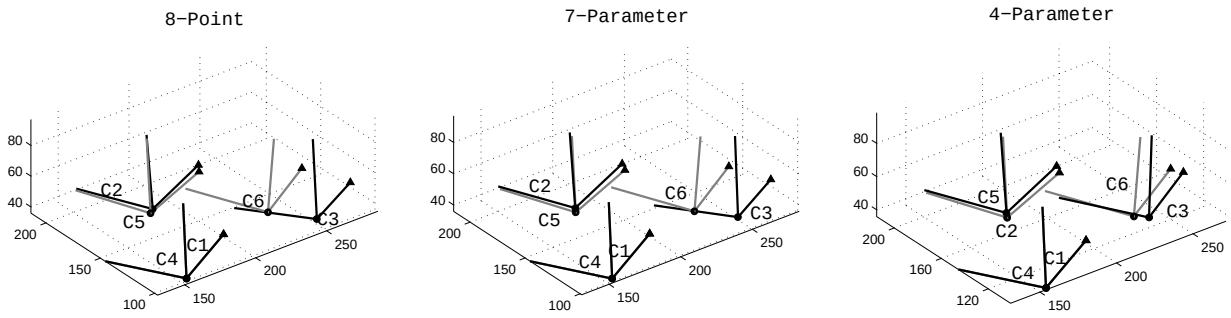


Figure 5-33: *Solutions to camera rotations and translations superimposed onto ground truth for the stable parameterization of the* `knossos` *simulation. Camera labels (1,2,3) are the estimates and camera labels (5,6,7) are the corresponding ground truth poses. Camera configurations are aligned with first view of ground truth. Camera poses are extracted from pairwise fundamental matrices computed by (a) 8-point algorithm, (b) nonlinear 7-parameter optimization for each independent camera pair, (c) minimal parameterization: 7-parameter optimization for adjacent views, 4-parameter constrained optimization for views 1 and 3.*

Chapter 6

Discussion

This final chapter discusses several important directions for the study of minimally parameterized projective models of multiple view geometry.

6.1 Future Directions

6.1.1 Robustness and Degeneracy

A key issue in modeling any view configuration with a fixed geometric model is the joint problem of image point robustness and model degeneracy. Robust methods, such as the RANSAC algorithm used in this work, attempt to delete outliers from a set of image point matches and fit a geometric model using only a subset of data points that provide a good fit. Model degeneracy occurs when the chosen image point set does not sufficiently constrain the geometric model, so many inaccurate estimates might explain the data equally well. Torr treats these two issues simultaneously for the problem of 3D camera analysis by designing methods that automatically determine a robust subset of point matches while detecting degeneracies in the camera configurations [42, 44].

These methods are ultimately necessary for applying the minimally parameterized models described in this work to the general analysis of image sequences. First, robustness measures are useful for determining which view pairs have sufficient reliable point matches to guarantee accurate independent estimates of their epipolar relations and which pairs do not. Those view pairs whose correspondences have too many outliers are good candidates

for the trifocally constrained epipolar parameter estimation.

Second, degeneracy detection for view triplet configurations is important for deciding which views may be used to constrain others. For instance, three collinear views will not have stable trifocal line estimates, so such views should not be selected for enforcing the general trifocal constraints on epipolar geometries. Even when three views are not collinear, the special cases of pure rotation, pure translation, or true affine epipolar geometry should be detected and handled by a simplified three-view model.

6.1.2 Linking Models

Several interesting questions arise from considering degenerate camera motions and point configurations. For instance, if the only reliable point matches between two view pairs are coplanar in the scene, there is insufficient data to compute the epipolar geometry. One question to ask is whether the trifocal constraints from other views provide enough additional information to determine the fundamental matrix from the coplanar scene points. It may be that the elegant relations between the homographies relating two views of coplanar scene points and the fundamental matrices may be exploited [27].

More generally, if a global minimal model of N views is desired, it is necessary to establish the dependencies among N epipolar geometries. For instance, if all consecutive triplets in an image sequence are trifocally constrained, that is, if the epipolar geometry between each first and third view pair in a triplet is trifocally constrained using the middle view, then a total of $7(N-1)+4(N-2) = 11N-15$ parameters have been used, 7 parameters for each of the $N-1$ consecutive pairs, and 4 parameters for each of the non-consecutive pairs in the $N-2$ triplets. In practice, it may not be the case that consecutive views in a collection have reliably matched points, so the partition of epipolar parameters into independent and dependent sets will be more complex.

These topics are important for future work in understanding how minimally parameterized projective models can best be leveraged to estimate global camera geometry from arbitrarily long uncalibrated image sequences.

6.2 Summary

This thesis has addressed the general problem of how to find globally consistent and accurate estimates of multiple-view camera geometry from uncalibrated imagery of an extended scene. For three uncalibrated views in general position, we have presented a new, minimally parameterized projective model of the trifocal geometry when the optical centers of the cameras are in general position. We have derived the geometric constraints on the three pairwise epipolar geometries in a view triplet, and constructed a representation that guarantees agreement between all pairwise relations.

To test this model, a nonlinear optimization algorithm for estimating the minimally parameterized model was designed and tested on synthetic and real image sequences. The estimation procedure takes as input three sets of matching point pairs, one for each view pair, and returns a set of dependent epipolar parameters that correspond to a unique projective camera configuration. The results are tested on commonly occurring practical scenarios in which one view pair in a triplet has less view overlap and therefore fewer point matches than the other two pairs. The results show that the trifocally constrained epipolar geometry improves the accuracy of view pairs with fewer point matches.

The numerically stability of the minimally parameterized model estimation procedure is tested by analyzing the local shape of the objective function in the parameter space evaluated over matching image points. The lower dimensionality of the trifocally constrained epipolar parameter space allows optimization procedures to settle at a more accurate global optimum than the unconstrained parameter space. In addition, critical camera configurations are shown to be unstable under certain parameterizations. A new, generalized parameterization that adapts the representation of the epipolar and trifocal geometry to the current estimate is presented and tested on synthetic and real uncalibrated imagery. The results show that the objective function of the generalized parameterization is numerically stable, and the trifocally constrained estimation improves the accuracy of geometry estimates.

Bibliography

[1] M. Antone and S. Teller. Scalable, Controlled Imagery Capture in Urban Environments. Technical Report 825, MIT LCS, September 2001.

[2] Adrien Bartoli and Peter Sturm. Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters. Technical Report RR-4236, INRIA, August 2001.

[3] Adrien Bartoli, Peter Sturm, and Radu Horaud. A Projective Framework for Structure and Motion Recovery from Two Views of a Piecewise Planar Scene. Technical Report RR-4970, INRIA, October 2000.

[4] R.C. Brown. The Bundle Adjustment. *Int. Archives in Photogrammetry*, 21(3), 1976.

[5] Gabriella Csurka, Cyril Zeller, Zhengyou Zhang, and Olivier D. Faugeras. Characterizing the Uncertainty of the Fundamental Matrix. *Computer Vision and Image Understanding*, 68(1):18–36, October 1997.

[6] B. Faugeras, O. D.and Mourrain. On the Geometry and Algebra of the Point and Line Correspondences Between N Images. In *International Conference on Computer Vision*, pages 951–956, 1995.

[7] O. D. Faugeras. What Can Be Seen In Three Dimensions With an Uncalibrated Stereo Rig? In *European Conference on Computer Vision, LNCS 588*, pages 563–578, 1992.

[8] O. D. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.

[9] O. D. Faugeras and Q. T. Luong. *The Geometry of Multiple Images*. MIT Press, 2001.

[10] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. Camera Self-Calibration: Theory And Experiments. In *European Conference on Computer Vision, LNCS 588*, pages 321–334, 1992.

[11] O. D. Faugeras and S. J. Maybank. Motion from Point Matches: Multiplicity of Solutions. In *International Conference on Computer Vision*, volume 4, pages 225–246, 1990.

[12] O. D. Faugeras and T. Papadopoulo. Grassmann-Cayley Algebra for Modeling Systems of Cameras and the Algebraic Equations of the Manifold of Trifocal Tensors. Technical Report RR-3225, INRIA, Sophia Antipolis, France, July 1997.

[13] O. D. Faugeras and T. Papadopoulo. A Nonlinear Method for Estimating the Projective Geometry of Three Views. In *International Conference on Computer Vision*, pages 477–484, 1998.

[14] C. Fermuller and Y. Aloimonos. Observability of 3D Motion. *International Journal of Computer Vision*, 37(1):43–62, June 2000.

[15] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24:381–385, 1981.

[16] A. W. Fitzgibbon and A. Zisserman. Automatic Camera Recovery for Closed or Open Image Sequences. In *European Conference on Computer Vision*, volume 1, pages 311–326, 1998.

[17] R. Hartley. Estimation of Relative Camera Positions for Uncalibrated Cameras. In *European Conference on Computer Vision*, pages 579–587, 1992.

[18] R. Hartley. Lines and Points in Three Views: A Unified Approach. In *ARPA Image Understanding Workshop*, pages II:1009–1016, 1994.

[19] R. Hartley. In Defence of the Eight-Point Algorithm. In *International Conference on Computer Vision*, pages 1064–1070, 1995.

[20] R. Hartley and P. Sturm. Triangulation. In *ARPA Image Understanding Workshop*, pages 957–966, November 1994.

[21] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[22] A. Heyden. A Common Framework for Multiple View Tensors. In *European Conference on Computer Vision*, pages 3–19, 1998.

[23] K. Kanatani. Model selection for geometric fitting: Geometric AIC and Geometric MDL. Technical Report 1, Memoirs of the Faculty of Engineering, Okayama University, December 2001.

[24] S. Laveau and O. D. Faugeras. 3-D Scene Representation as a Collection of Images and Fundamental Matrices. Technical Report RR-2205, INRIA, February 1994.

[25] L. C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from Two Projections. In *Nature*, volume 293, pages 133–135, 1981.

[26] M. Lourakis and R. Deriche. Camera Self-Calibration Using the Singular Value Decomposition of the Fundamental Matrix: From Point Correspondences To 3D Measurements. Technical Report RR-3748, INRIA, August 1999.

[27] Q. T. Luong and O. D. Faugeras. Determining the Fundamental Matrix with Planes: Instability and New Algorithms. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 489–494, 1993.

[28] Q. T. Luong and T. Viéville. Canonical Representations for the Geometries of Multiple Projective Views. *Computer Vision and Image Understanding*, 64(2):193–229, September 1996.

[29] Y. Ma, J. Kosecka, and S. Sastry. Optimization Criteria, Sensitivity and Robustness of Motion and Structure Estimation. *International Journal of Computer Vision*, 44(3):219–249, 2001.

[30] D. Marr and T. Poggio. Cooperative Computation of Stereo Disparity. *Science*, 194:282–287, 1976.

[31] B. K. P. Horn. Recovering Baseline and Orientation from Essential Matrix. `http://www.ai.mit.edu/people/bkph/papers/essential.pdf`, 1990.

[32] B. K. P. Horn. Projective Geometry Considered Harmful. `http://www.ai.mit.edu/people/bkph/papers/harumful.pdf`, 1999.

[33] J. Oliensis. The Error Surface for Structure from Motion. Technical report, NECI, 2001.

[34] C. Rothwell, O. D. Faugeras, and G. Csurka. A Comparison of Projective Reconstruction Methods for Pairs of Views. *Computer Vision and Image Understanding*, 68(1):37–58, 1997.

[35] L. S. Shapiro, A. Zisserman, and M. Brady. Motion from Point Matches Using Affine Epipolar Geometry. In *European Conference on Computer Vision, LNCS 801*, pages 73–84, 1994.

[36] A. Shashua. Trilinearity in Visual Recognition by Alignment. In *European Conference on Computer Vision, LNCS 800*, pages 479–484, 1994.

[37] A. Shashua. Algebraic Functions for Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, August 1995.

[38] M. E. Spetsakis and Y. Aloimonos. A Unified Theory of Structure from Motion. In *DARPA90*, pages 271–283, 1990.

[39] G. P. Stein. Lens Distortion Calibration Using Point Correspondences. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 602–609, 1997.

[40] P. Sturm and B. Triggs. A Factorization Based Algorithm for Multi-image Projective Structure and Motion. In *European Conference on Computer Vision*, pages II:709–720, 1996.

[41] C. Tomasi. Shape and Motion from Image Streams: A Factorization Method. Technical Report CMU-CS-91-172, CMU, September 1991.

[42] P. H. S. Torr, A. Fitzgibbon, and A. Zisserman. The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences. *International Journal of Computer Vision*, 32(1):27–44, August 1999.

[43] P. H. S. Torr and A. Zisserman. Robust Parameterization and Computation of the Trifocal Tensor. In *Image and Vision Computing*, pages 591–605, 1997.

[44] P. H. S. Torr, A. Zisserman, and S. J. Maybank. Robust Detection of Degenerate Configurations Whilst Estimating the Fundamental Matrix. *Computer Vision, Graphics, and Image Processing*, 71:312–333, 1998.

[45] B. Triggs. Matching Constraints and the Joint Image. In *International Conference on Computer Vision*, pages 338–343, 1995.

[46] B. Triggs. Factorization Methods for Projective Structure and Motion. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 845–851, 1996.

[47] B. Triggs. Optimal Estimation of Matching Constraints. In R. Koch and L. Van Gool, editors, *European Workshop on 3D Structure from Multiple Images of Large-scale Environments, SMILE '98*. Springer Verlag LNCS, 1998.

[48] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle Adjustment—A Modern Synthesis. In *International Workshop on Vision Algorithms: Theory and Practice*, pages 298–372, 1999.

[49] R. Y. Tsai and T. S. Huang. Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):13–27, January 1984.

[50] S. Ullman. *The Interpretation of Visual Motion*. MIT Press, 1979.

[51] C. Zeller and O. D. Faugeras. Camera Self-Calibration from Video Sequences: The Kruppa Equations Revisited. Technical Report RR-2793, INRIA, 1996.

[52] Z. Y. Zhang. A New Multistage Approach to Motion and Structure Estimation: From Essential Parameters to Euclidean Motion Via Fundamental Matrix. Technical Report RR-2910, INRIA, 1996.

[53] Z. Y. Zhang. Determining the Epipolar Geometry and its Uncertainty: A Review. *International Journal of Computer Vision*, 27(2):161–195, March 1998.

[54] Z. Y. Zhang. Understanding the Relationship Between the Optimization Criteria in Two-View Motion Analysis. In *International Conference on Computer Vision*, pages 772–777, 1998.

[55] Z. Y. Zhang, R. Deriche, O. D. Faugeras, and Q. T. Luong. A Robust Tecnhique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. *Artificial Intelligence Journal*, 78(1-2):87–119, October 1995.